



Calhoun: The NPS Institutional Archive
DSpace Repository

Theses and Dissertations

1. Thesis and Dissertation Collection, all items

2017-12

Spatial-temporal co-occurrence in the maritime domain

Sollish, Andrew B.

Monterey, California: Naval Postgraduate School

<http://hdl.handle.net/10945/56811>

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun



<http://www.nps.edu/library>

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

SPATIAL-TEMPORAL CO-OCCURRENCE IN THE MARITIME DOMAIN

by

Andrew B. Sollish

December 2017

Thesis Advisor:

Co-Advisor:

Co-Advisor:

Sean F. Everton

Neil C. Rowe

Wayne Porter

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE December 2017	3. REPORT TYPE AND DATES COVERED Master's thesis		
4. TITLE AND SUBTITLE SPATIAL-TEMPORAL CO-OCCURRENCE IN THE MARITIME DOMAIN			5. FUNDING NUMBERS	
6. AUTHOR(S) Andrew B. Sollish				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Seventh Fleet; NPS Naval Research Program			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB number ____N/A____.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (maximum 200 words) To prevent and stop illegal and malicious maritime activity, it is necessary to better understand the people, places, organizations, and vessels that contribute to those activities. This thesis examines whether spatial-temporal co-occurrence is useful for finding connections among vessels and providing understanding of their activities. A spatial-temporal co-occurrence means that two or more entities can be found nearby at the same time. To identify those occasions, we used two datasets, Automatic Identification System (AIS) records of vessel positions and time and company information on the ownership of vessels (used to confirm if vessels were affiliated). Three experiments studied co-occurrence among groups of vessels including randomly selected vessels, vessels of interest, and locations of interest. Vessels were identified for each of the three experiments and their co-occurrences were identified. We found that co-occurrence manifested in different ways and it was not especially useful in demonstrating affiliation among vessels.				
14. SUBJECT TERMS maritime, co-occurrence, spatial, temporal, automatic identification system			15. NUMBER OF PAGES 73	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UU	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

SPATIAL-TEMPORAL CO-OCCURRENCE IN THE MARITIME DOMAIN

Andrew B. Sollish
Lieutenant, United States Navy
B.A., University of Utah, 2008

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN COMPUTER SCIENCE
and
MASTER OF SCIENCE IN DEFENSE ANALYSIS**

from the

**NAVAL POSTGRADUATE SCHOOL
December 2017**

Approved by: Sean F. Everton
Thesis Advisor

Neil C. Rowe
Co-Advisor

Wayne Porter
Co-Advisor

John Arquilla
Chair, Department of Defense Analysis

Peter J. Denning
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

To prevent and stop illegal and malicious maritime activity, it is necessary to better understand the people, places, organizations, and vessels that contribute to those activities. This thesis examines whether spatial-temporal co-occurrence is useful for finding connections among vessels and providing understanding of their activities. A spatial-temporal co-occurrence means that two or more entities can be found nearby at the same time. To identify those occasions, we used two datasets, Automatic Identification System (AIS) records of vessel positions and time and company information on the ownership of vessels (used to confirm if vessels were affiliated). Three experiments studied co-occurrence among groups of vessels including randomly selected vessels, vessels of interest, and locations of interest. Vessels were identified for each of the three experiments and their co-occurrences were identified. We found that co-occurrence manifested in different ways and it was not especially useful in demonstrating affiliation among vessels.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
II.	RELATED WORK	3
A.	MARITIME DOMAIN AWARENESS	3
B.	SPATIAL-TEMPORAL ANALYSIS.....	4
1.	Cellular.....	4
2.	Animal Networks	5
3.	Web, Email, and Social Media.....	6
III.	METHODS	9
A.	FINDING CO-OCCURRENCES	9
1.	Experiment I: Randomly Selected Vessels	9
2.	Experiment II: Locations of Interest (LOI).....	10
3.	Experiment III: Vessels of Interest (VOI)	12
4.	Confirming Social Connection.....	13
B.	THE AIS DATA	13
C.	AIS TECHNOLOGY	15
1.	An Overview	15
2.	AIS Vulnerabilities.....	16
D.	MERGING AND CLUSTERING OWNERSHIP DATA	18
E.	REPRESENTING CO-OCCURRENCES.....	21
IV.	RESULTS	25
A.	EXPERIMENT I: RANDOMLY SELECTED VESSELS.....	25
B.	EXPERIMENT II: LOCATIONS OF INTEREST (LOI)	30
C.	EXPERIMENT III: VESSELS OF INTEREST (VOI)	41
D.	NOTABLE ASSOCIATIONS.....	47
V.	CONCLUSION	51
A.	FINDINGS.....	51
B.	FURTHER RESEARCH.....	52
	LIST OF REFERENCES.....	53
	INITIAL DISTRIBUTION LIST	55

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 1.	Location of Interest Hong Kong. Adapted from [13].	10
Figure 2.	Location of Interest Spratly. Adapted from [14].	11
Figure 3.	Location of Interest Spratly. Adapted from [15].	12
Figure 4.	Area of all AIS Records. Adapted from [18].	14
Figure 5.	Ship Length Histogram	16
Figure 6.	Ship Type Histogram	17
Figure 7.	Itu Aba Island – 7.79 Km Radius. Adapted from [23].	22
Figure 8.	Itu Aba Island (Zoomed In). Adapted from [24].	22
Figure 9.	Co-occurrence Scatter Plot (Random Vessels) ($r = 0.68$)	26
Figure 10.	Number of Locations Histogram (Random)	27
Figure 11.	Total Co-occurrences Histogram (Random)	28
Figure 12.	ROC Unique Locations (Random)	29
Figure 13.	ROC Total Co-occurrences (Random)	30
Figure 14.	Co-occurrence Scatter Plot (Hong Kong) ($r = 0.8$)	31
Figure 15.	Co-occurrence Scatter Plot (Spratly) ($r = 0.07$)	32
Figure 16.	Unique Locations Histogram (Hong Kong)	33
Figure 17.	Unique Locations Histogram (Spratly)	34
Figure 18.	Total Co-occurrences Histogram (Hong Kong)	35
Figure 19.	Total Co-occurrences Histogram (Spratly)	36
Figure 20.	Unique Locations ROC (Hong Kong) ($AUC = 0.78$)	37
Figure 21.	Unique Locations ROC (Spratly) ($AUC = 0.52$)	38
Figure 22.	Total Co-occurrences ROC (Hong Kong) ($AUC = 0.77$)	39
Figure 23.	Total Co-occurrences ROC (Spratly) ($AUC = 0.64$)	40

Figure 24.	Co-occurrence Scatter Plot (COSCO) ($r = 0.9$)	41
Figure 25.	Unique Locations Histogram (COSCO)	42
Figure 26.	Total Co-occurrences Histogram (COSCO)	43
Figure 27.	Unique Locations ROC (COSCO) ($AUC = 0.51$)	44
Figure 28.	Total Co-occurrences ROC (COSCO) ($AUC = 0.51$)	45
Figure 29.	Unique Locations ROC (All Samples)	46
Figure 30.	Total Co-occurrences ROC (All Samples)	47

LIST OF TABLES

Table 1.	Location of Interest—Spratly Island Locations.....	12
Table 2.	Generic Terms Example	18
Table 3.	Different Postal Code Example	19
Table 4.	Different Floor Example	19
Table 5.	Missing Building or Unit Number Example	19
Table 6.	Affiliates Example	20
Table 7.	Affiliates Example	20
Table 8.	Affiliates Example	21
Table 9.	Interesting Pairs (Random Samples).....	48
Table 10.	Interesting Pairs (Hong Kong)	48
Table 11.	Interesting Pairs (Spratly)	49
Table 12.	Interesting Pairs (COSCO)	49

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AIS	automatic identification system
AUC	area under curve
COSCO	China Ocean Shipping Company
IMO	International Maritime Organization
EBM	entropy based model
GEOSO	geospatial social model
GPS	global positioning system
LOI	location of interest
MDA	maritime domain awareness
MMSI	maritime mobile service identity
MOU	memorandum of understanding
ROC	receiver operator characteristic
SNA	social network analysis
SOE	state-owned enterprise
UTC	coordinated universal time
VOI	vessel of interest

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank Rob Schroeder for our many conversations about the research project and for his web-scraping R code. I would like to thank Dr. Mikhail Auguston for his suggestions on algorithms and time complexity. I would also like to thank Dr. John Mittleman for his work on pre-processing the AIS data.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The maritime domain is the major way that freight is transported around the world. Approximately 80% of global trade by volume is conducted by the shipping industry [1]. Unfortunately, some commerce via ship includes illegal activity in the form of piracy, trafficking (drugs, munitions, or people), exploitation of marine resources, violations of environmental law, and other nefarious activity. Complicating matters are economic and political conflicts that occur inside waters where jurisdiction is highly contentious, e.g., the Spratly Islands in the South China Sea.

Due to the amount of maritime commerce and the volume of illegal and contentious activity, there is a need to improve maritime domain awareness (MDA) for grey zone activity (activity that occurs on the spectrum between peacetime and conflict). To prevent and correct malicious activity, it is necessary to better understand the people, places, organizations, and vessels that contribute to those activities.

One way to improve maritime domain awareness is via the field of social network analysis (SNA). SNA is a set of theories, techniques, and metrics, borrowed from various disciplines, that are used to study underlying social structures. It is not to be confused with link analysis, the study of entities connected to other entities. SNA goes a step further, attempting to understand what underlies those connections. It has been successfully applied to social structures such as religious organizations, companies, and universities, as well as illicit social structures like crime syndicates and terrorist groups.

However, before SNA can be applied to a problem, a network graph must be constructed. This is a process of identifying the entities or nodes and the relational connections or edges between them. In the maritime domain, nodes could be vessels, companies, individuals, and ports. At present, there are no easy ways to populate network graphs in the maritime domain, especially networks that prefer to go undiscovered. This thesis suggests spatial-temporal co-occurrence as a way of finding relational connections in a network graph. “Spatial-temporal co-occurrence” means that two or more entities can be found nearby at the same time. How one defines “nearby” and “the same time” is

circumstantial. Defining the necessary spatial and temporal thresholds is an inexact science and it will depend upon the problem in which they are applied. In the maritime domain, this could mean two ships that report their position within a mile within several hours.

For people, even a small number of co-occurrences between them results in a significantly greater likelihood of social connection [2]. Some work conducted on the idea of co-occurrence is discussed in Chapter II.

To test co-occurrence in the maritime domain, this thesis studied two datasets. The first dataset was historical Automatic Identification System (AIS) records which make use of on-board vessel transponders that periodically report position and time. The second dataset was vessel ownership records. The AIS data were used to find co-occurrences among vessels, and the ownership records served as a ground truth to confirm whether vessels were affiliated. The two datasets and the methods of testing co-occurrence are discussed in greater detail in Chapter III.

Chapter II discusses related work on co-occurrence applied to areas outside the maritime domain. There is also discussion of AIS's application in the maritime domain. In Chapter III the AIS system, the data studied for this research, and the methods applied are all discussed in greater detail. In Chapter IV, the findings are presented and discussed. Lastly, in Chapter V are the conclusions and research suggested for further study.

II. RELATED WORK

Although real-time ship data is a relatively new technology, and there is a growing body of research using it, there has been only a limited amount of work analyzing and exploiting this data for in-depth maritime domain awareness and supporting networks. Spatial-temporal analysis is a natural fit for the ship data due to its geographical and temporal components. Common research for it includes surveillance, anomaly detection, clustering, and course prediction. Nonetheless, revealing maritime networks from such data and exploring their spatial-temporal co-occurrence phenomena is untested.

Identifying social networks based upon spatial-temporal co-occurrence of events and relationships has been explored in other realms. The rapid proliferation of mobile phones has provided rich databases with which to mine co-occurrences. There is also a growing body of spatial-temporal research studying social-media websites that offer extensive amounts of data. Many social-media sites offer the user the ability to tag themselves in space and time, post pictures and video, as well as to connect with people. Spatial-temporal co-occurrence is also being explored in animal-network research. Monitoring groups of animals and their environments makes understanding social networks within groups of animals easier and more accurate.

What follows is first a survey of work on the use of ship data, followed by spatial-temporal research in fields outside of the maritime domain.

A. MARITIME DOMAIN AWARENESS

Real-time ship data can provide a rich source of insight regarding historical and present maritime activity and relationships. Ray, Grancher, Thibaud, and Etienne [3] presents a proof-of-concept for an expert system designed to improve maritime domain awareness. They suggest a software architecture that takes in dynamic data and uses rules to produce notifications of particular maritime activity. That activity could be anomalous behavior or notifications when vessels have entered a restricted area.

Using such data to improve maritime domain awareness has limitations. Last, Bahlke, Hering-Bertram, and Linsen [4] reviewed an extensive amount of data in analyzing vessel-prediction algorithms. They found that prediction solely using this data is difficult for several reasons. One is that there currently is a lack of consistency in short-time reporting intervals. This impacts the timeliness of data fields such as, rate of turn and heading which are important in predicting vessel course changes. Other problems are related to incorrect transmitter placement, network overload, land obstruction, and data loss over VHF. This leads the authors to conclude that real-time ship data should be supplemented with additional data sources.

B. SPATIAL-TEMPORAL ANALYSIS

1. Cellular

One way of identifying and studying co-occurrences among people is via cell phones. Eagle and Pentland [5] used data collected from 100 mobile phones on the campus of MIT. They found that spatial-temporal co-occurrence provided insights into social relationships. They designed a Gaussian mixture model to correlate proximity patterns with the types of relationship. Using surveys, subjects were asked to name the people with whom they spent time in and out of the workplace and who they consider close friends. They could reach 90% accuracy in distinguishing between colleagues, friends, and “inner-circle” friends using only data from the mobile phones, as confirmed via survey. Eagle and Pentland further suggest that including communication logs and using more powerful modeling techniques like support vector machines could increase accuracy.

Eagle, Pentland, and Lazer [6] inferred friendship by classifying proximity by context. For example, they distinguished proximity events between campus and off-campus, daytime and nighttime, and weekday and weekend. This tested the hypothesis that co-occurrence of people at night and over the weekend at an off-campus locale is a better indication of friendship than co-occurrence during the day in a weekday on campus. They accurately predicted 96% of symmetric reports of non-friendship and 95%

of reports of symmetric friendship. Also, in [5] and [6] mobile-phone data is shown to be highly effective at predicting social relationships.

Wang, Pedreschi, Song, and Giannotti [7] analyzed the communication records of six million mobile-phone users. They claimed a correlation between the similarity of users and their closeness within a social network. They proposed three categories to make correlations. The first is “mobile homophily,” the distance of the nearest cell phone tower for each placed call. The second is social connectedness, which they measure in a variety of ways including the Jaccard coefficient, Katz, common neighbors, and Adamic-Adar. The ways compare two sets; in this case, they compared the number of neighbors that two nodes have in common, provided they have at least one cell-phone call between them. The third is the number of calls placed between two users. Wang et al. found that these co-location measures provided for highly accurate link prediction. The authors define link prediction to mean detecting, among all the pairs of nodes that did not call one another in the past, those pairs that will communicate with each other in the future. Additionally, they found that the combination of mobility measures with pre-existing network-closeness measures can significantly improve link prediction in a supervised classification.

2. Animal Networks

Krause et al. [8] identified a variety of methods for studying animal networks, what they call reality mining. It is a way of studying social behavior using sensor data rather than by direct observation or survey. Such sensor data can be generated by the Global Positioning System (GPS), cameras, accelerometers, radar, or any technology able to geo-locate an entity in space and time.

The authors note that large amounts of dynamic data can allow researchers to better understand animal social structure. They suggest that social-network metrics and hidden Markov models can help reveal routines and behavior that precede certain states. The process of reality mining and the subsequent analytical techniques could be analogous in their application to the maritime environment where ships, crews, and owners are woven into a social network.

Using sensor technologies and spatial-temporal analysis is not without its pitfalls. Psorkis et al. [9] note some of the difficulties of doing such research on ecological systems. They studied thousands of birds marked with transponders on a grid of sensor locations with 16 bird feeders. Their goal was to discover the social structure of the birds by measuring the co-occurrence of birds at various feeders. They found difficulty in deciding on appropriate time windows that would yield meaningful social associations. Time windows too small would miss important co-occurrences, while windows too large would lead to an overrepresentation of social connections. It becomes a precision/recall problem of finding the right balance.

3. Web, Email, and Social Media

Increasingly, social scientists are studying social media. In studies involving spatial-temporal co-occurrence, social media not only provide a source of spatial and temporal data, but also provide a method to confirm ties identified or expected by other means. Social media also offers researchers the ability to easily mine data with millions of nodes and edges.

For example, Pham, Shahabi, and Hu [10] give a model to measure co-occurrence in space and time. Their geospatial social model (GEOSO) reflects the value of co-occurrence spread over multiple locations and co-occurrence of the same amount or greater but in fewer locations. For instance, if two pairs of users each have six co-occurrence events, but the first pair has all six co-occurrences in the same place, whereas the second has them in six different locations, the model asserts that the second pair has a stronger social connection. They measure the strength of a connection as a Euclidean distance. They create a master vector that reflects the maximum value of co-occurrence a pair of users can have in a chosen space-time grid. They measure the strength of connection between two users as the inner product of the co-occurrence vector and the master vector.

These ideas are further explored in [11], which uses data from a social-media application called GOWALA that allows users to share their location via “check-ins.” Their entropy based model (EDM) predicts the strength of social connections as a

function of both the diversity of co-occurrence and the weighted frequency of co-occurrence. Weighted frequency is designed to give greater significance to co-occurrence in uncommon space and time.

Crandall et al. [2] studied the social-media application Flickr that allows users to share photos and video in addition to connecting with one another. They used the geo-tag feature of photos posted on Flickr, and if two users posted photos of the same place at relatively the same time, they suspected a social connection. The reported connections on Flickr, in the form of friends/family lists, served as confirmation of a suspected social connection. An important takeaway of their research was that as cell size, the area where co-occurrence would be observed, decreases, the probability of social connection does not always improve. However, co-occurrence in a cell the size of a home or small business is greater cause for social connection than co-occurrence in a cell the size of Mexico. They did find, like [10] and [11], that co-occurrence in a diverse set of locations was a stronger indicator of friendship.

Lauw, Lim, Pang, and Tan [12] discussed the size of spatial and temporal cells and proposed formulas for the precision of spatial and temporal co-occurrences. Co-occurrences during unusual time periods and co-occurrences in rarer locales provide greater precision, translating to better scores in their proposed formulas. They used webpage requests from wireless computer users on their university campus and hypothesized that the requests for the same page indicate shared interests or collaboration. They used demographic similarity to affirm or reject such connections; co-occurrence events were cross-referenced with demographic data of the two users including academic major, year of graduation, and status (graduate, undergraduate, etc.). Ultimately, their approach is a test of homophily, the idea that people who are similar tend to associate with one another. The idea of using demographic data as a means of confirming social connections is discussed in greater detail in Chapter III.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODS

A. FINDING CO-OCCURRENCES

Co-occurrence can manifest itself in different ways depending on the situation. For instance, vessels that collaborate on oceanographic or geologic engineering projects at sea show different patterns of co-occurrence than vessels that have competing interests in a certain area. A group of shrimp boats may each have different owners, travel from the same port or nearby ports, and visit the same areas at different times and without being in close proximity. On the other hand, vessels collaborating to build a canal may be of different types, travel from different ports, but appear together in close proximity. To test whether co-occurrence helps to identify affiliations among vessels, it is helpful to explore various scenarios.

Three experiments were used to explore the idea of maritime co-occurrence. First, vessels were selected randomly from the dataset and tested for differences. Second, location(s) of interest (LOI) were selected and the vessels that were reported at those locations were identified, then co-occurrences among the vessels were retrieved. Third, vessel(s) of interest (VOI) were identified as those from a maritime conglomerate company with multiple subsidiaries and many vessels under ownership.

1. Experiment I: Randomly Selected Vessels

In the first experiment, randomly selected vessels served as a control group for identifying relationships. There were five iterations of randomly selected vessels that yielded between 127–175 different vessels in each of the five groups, though the same vessels could appear in multiple groups. The motivation for selecting five separate iterations was to control for differences among the vessels. Having a mix of different types of vessels appearing in different geographic areas should improve the control group.

2. Experiment II: Locations of Interest (LOI)

Maritime domain awareness usually starts with either a location of interest or vessels of interest. Someone would start with a location of interest if, for example, the area has contentious sovereignty claims like islands in the South China Sea. Also, the area could be interesting if it is regulated for activities like commercial fishing or is a major chokepoint vulnerable to piracy. Therefore, there is a specific reason to explore whether co-occurrence is manifested in select locations of interest.

In this study, two locations were used as locations of interest. The first was at latitude 22.333 and longitude 114.120 in the vicinity of Hong Kong (Figure 1). This area was selected because it provided a coastal area with a high density of maritime traffic. The area for inclusion was a 7.79 KM radius around the center point. The rationale for selecting the radii of locations is given in Section E.

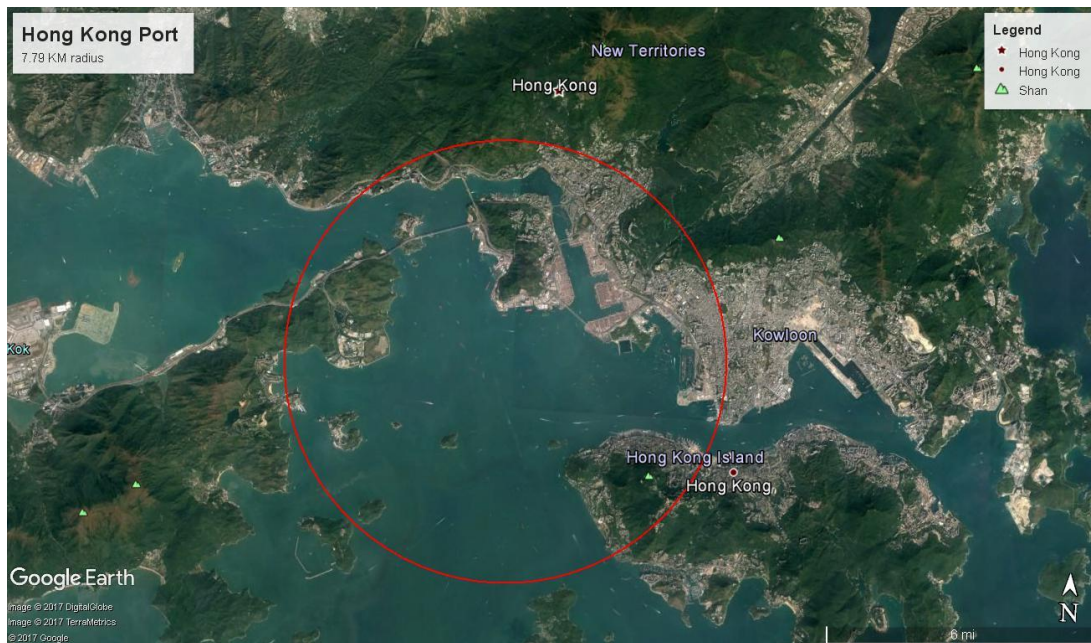


Figure 1. Location of Interest Hong Kong. Adapted from [13].

The second location of interest was a cluster of six islands in the South China Sea that are part of the Spratly Island chain (Figures 2 and 3). They are Itu Aba Island, Thitu

Island, West York Island, Spratly Island, Fiery Cross Reef, and Mischief Reef. Chinese vessels have regularly conducted reef enhancement activities there. The islands were considered as one location of interest due to their small size and close proximity. Table 1 lists their data.

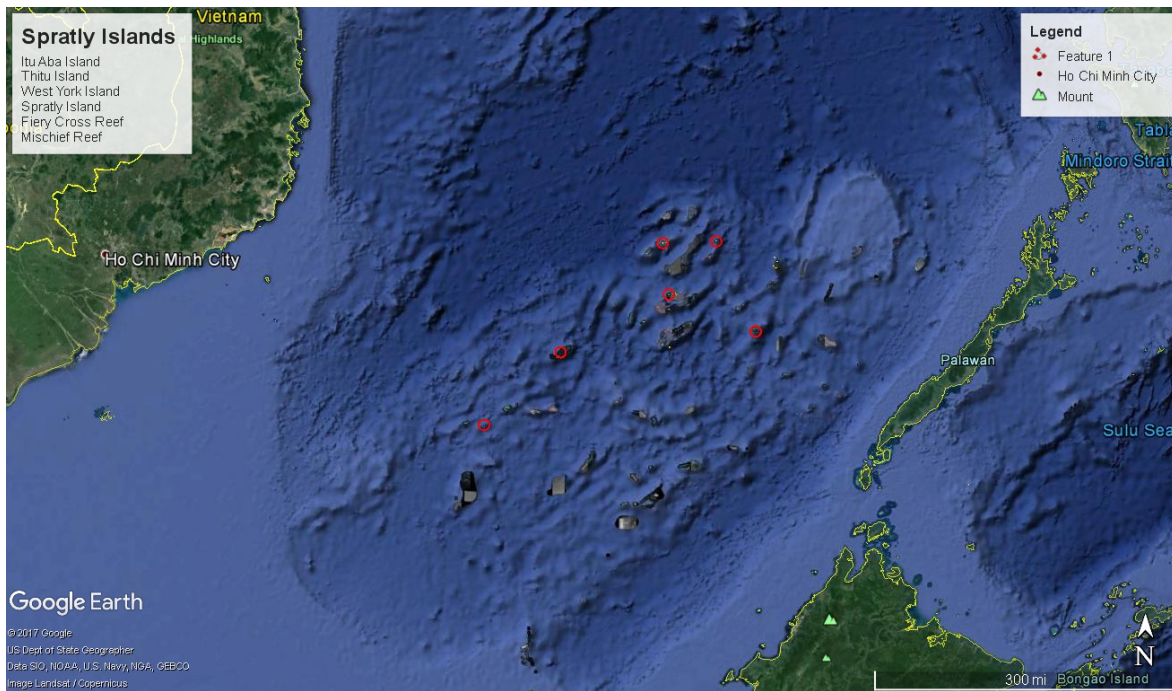


Figure 2. Location of Interest Spratly. Adapted from [14].



Figure 3. Location of Interest Spratly. Adapted from [15].

Table 1. Location of Interest—Spratly Island Locations

Name	Coordinates	Decimal Degrees	Area
Itu Aba Island	10°23'N 114°21'E	10.3750, 114.3699	46
Thitu Island	11°03'N 114°17'E	11.0521, 114.2904	37
West York Island	11°05'N 115°01'E	11.0780, 115.0022	18.6
Spratly Island	08°38'N 111°55'E	8.6423, 111.9243	13
Fiery Cross Reef	9°32'57"N 112°53'21"E	9.5482, 112.889167	Unknown
Mischief Reef	9°55'N 115°32'E	9.9013, 115.5354	Unknown

The traffic density and makeup is markedly different around the port of Hong Kong than the Spratly Islands. By studying two very different places, we could better determine if there are differences in the way that co-occurrence is manifested.

3. Experiment III: Vessels of Interest (VOI)

It is probably more common to have a predefined set of vessels of interest. For instance, one of the larger maritime companies in the world is China Ocean Shipping Company (COSCO). The company claims ownership of 1114 vessels and a dominance in

maritime logistics [16]. As a logistics company, most of the COSCO fleet is transport vessels like bulk carriers, tankers, and container vessels. However, the fleet also has specialized vehicles including semi-submersibles, heavy-lifters, and multi-purpose vessels. COSCO is a state-owned enterprise (SOE) with upwards of 300 different subsidiaries, many of which take on the COSCO name in some capacity, e.g. COSCO Pacific Ltd., or COSCO Shipping Lines. Using vessels owned by a conglomerate maritime company like COSCO provided a known maritime network, and its vessels could serve as vessels of interest.

4. Confirming Social Connection

After co-occurrences were identified among the vessels in each of the experiments, vessel ownership records were used as a ground truth to confirm whether the vessels were affiliated. That is, if a pair of vessels were owned by the same company, then they were considered affiliated. However, there are other forms of affiliation among vessels. For example, a vessel owned by a company that is a subsidiary of another, two vessels owned by different companies under the same contractual agreement, and several other situations could be considered to mean affiliation. Nonetheless, if co-occurrence can be proven useful for identifying vessel affiliation, then it can be applied to other situations in which ownership records are unavailable or inadequate to confirm affiliation. More about the techniques of identifying co-occurrence events is in Section E, and more on ownership records is in Section D.

B. THE AIS DATA

Our experiments used data from two sources. The AIS data were acquired from the Naval Research Laboratory. Because AIS records do not contain vessel ownership information, company data was acquired via web-scraping the Tokyo Memorandum of Understanding (MOU) [17]. Web scraping presented challenges for acquiring enough usable information. Queries of the Tokyo MOU were limited to 500 per 12-hour period, required a single vessel identifier like an International Maritime Organization (IMO) number, and could only be acquired one at a time. This made the process slow and

cumbersome. For future research, acquiring company information from a wider variety of maritime service companies should be preferred.

The AIS data encompassed a period between November 2014 and March 2016 and an area between Latitudes 4–40 and Longitudes 107–126 (Figure 4). This includes much of the Yellow Sea, East China Sea, South China Sea, and part of the Philippine Sea. The area serves as one of the busier maritime zones, and includes a significant amount of coastline as well as open ocean areas, economic resources (minerals, petroleum, fish stocks), and other features of special interest (reefs, islands). This data was large enough and diverse enough to conduct a thorough study.

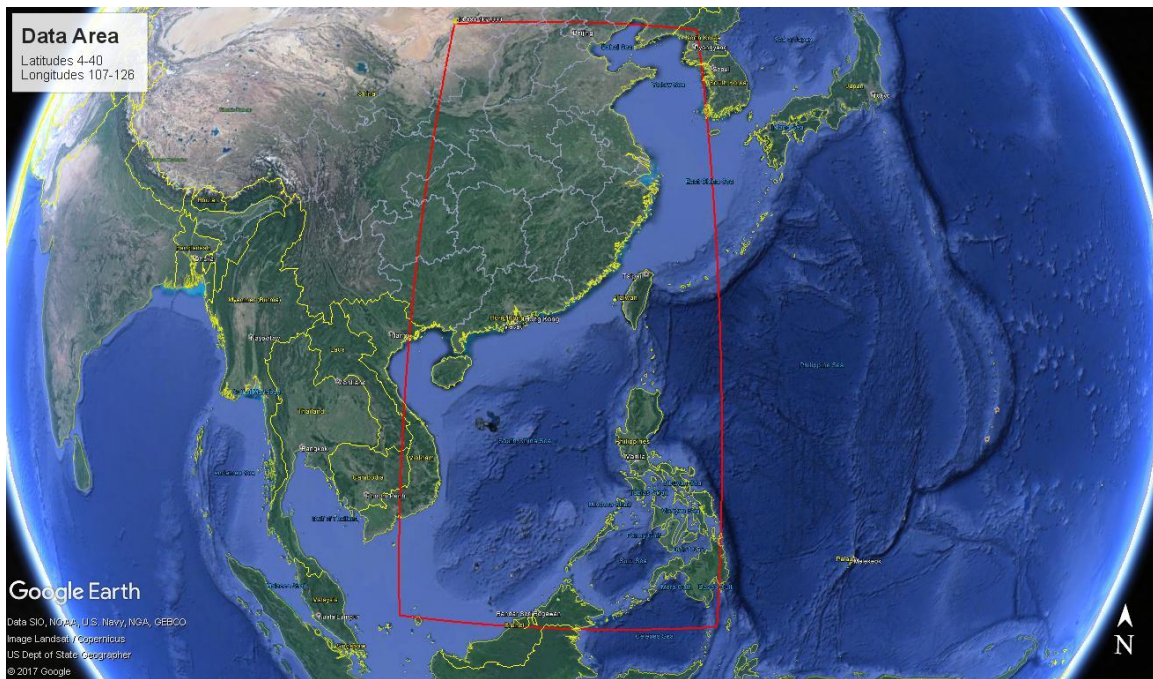


Figure 4. Area of all AIS Records. Adapted from [18].

The AIS data contained 5,976,086 records. Of all the records, 16% (983,912) contained missing or invalid IMO numbers. The two datasets, AIS and company data, only shared vessel IMO numbers, so it was only by these numbers that the two datasets could be joined. There were also invalid Maritime Mobile Service Identity (MMSI)

numbers that are unique nine-digit numbers used to temporarily identify ships. The AIS data also contained some apparently invalid ship names (e.g. “K9s(psid* 4u-pua@u K”), vessel speeds, and lengths. For example, some ships had lengths of zero, and some had recorded speeds well over 80 knots—an unachievable speed with existing technology.

C. AIS TECHNOLOGY

1. An Overview

The Automatic Identification System (AIS) is a transponder-based system for the maritime domain. Information generated from shipboard devices like GPS receivers and gyrocompasses is transmitted via an onboard VHF transceiver to coastal stations and AIS-equipped satellites. AIS data includes the MMSI number, the IMO which is a permanent seven-digit number associated with the ship hull, the name of the vessel, the call sign, the type of ship, the type of cargo, the ship dimensions, the draught, the destination, the estimated time of arrival, the current location, the time of the data in coordinated universal time (UTC) seconds, the current heading, the course, the speed, and the navigation status (i.e., underway, at anchor, etc.).

There are two classes of AIS transceivers, A and B. The A class is designed for larger commercial vessels and has more stringent performance and capability requirements. The equipment includes additional interfaces for measuring the rate of turn, the course, the compass heading and the GPS location, and the transmission periodicity is ten seconds to three minutes instead of the 30 seconds to three minutes for class B. The Class B transceivers are designed for smaller commercial and recreational vessels. The power and data-transmission requirements are more relaxed for class B than for class A.

The ability of vessels to transmit AIS data and see the data from other nearby vessels helps in collision-avoidance, its primary purpose. AIS can also be used to aid in navigation, track and monitor fleets, assist in maritime security, and help reveal vessel-network affiliations.

2. AIS Vulnerabilities

AIS equipment is not currently required on all vessels. Only vessels at least 300 gross tons and all passenger vessels are required to be equipped with AIS transponder equipment [19]. Smaller vessels and those of special types have different requirements depending upon the waters in which they operate. There has been a proliferation of AIS devices on vessels, but it has not yet been fully adopted.

Figure 5 shows the vessel lengths in the data. We can see that the listed vessels are predominantly over 50ft. long. In Figure 6, which shows vessel types, we can see that there are many more vessels of type 8 (tankers) than any other category.

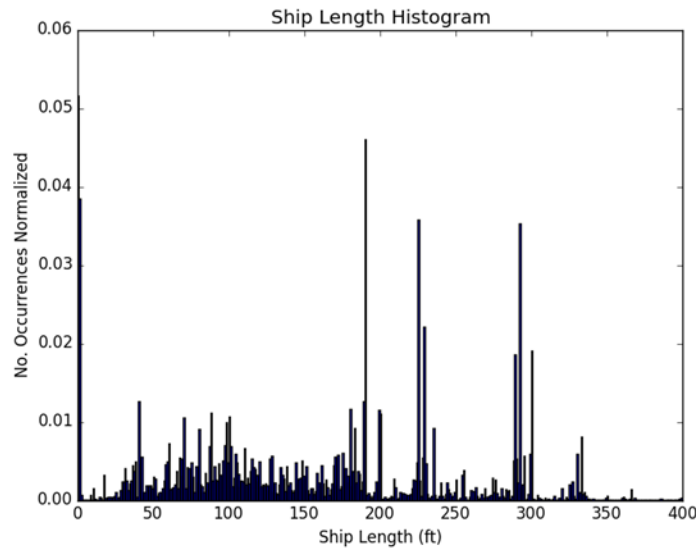


Figure 5. Ship Length Histogram

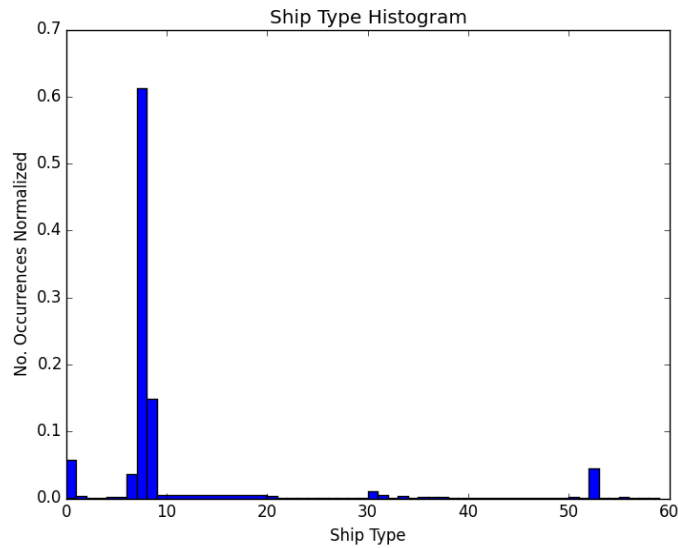


Figure 6. Ship Type Histogram

AIS technology has reliability problems. VHF transceivers require a line of sight and have a variable range depending on environmental factors and antenna placement. Using the more powerful class A transceiver, the horizontal range can be 20–30 miles [20]. The vertical range is much further but fewer satellites can receive vertical transmissions than there are coastal facilities able to receive line of sight (horizontal) transmissions. So vessels outside the range of ground-based stations have a greater likelihood of transmissions going un-received or dropped.

User error associated with inputting and transmitting AIS message traffic can also be a problem. Since certain information must be manually entered into AIS transponders, it is not uncommon to see erroneous MMSI numbers. A vessel could also broadcast the MMSI number of a different vessel to deliberately conceal its position or identity. The “Global Fishing Watch” claims they have found spoofing to occur less than 0.25% of the time [21]. Regardless, an analyst could be interested in the vessels conducting nefarious activity, and those vessels could be ones willing to spoof position or identity.

There are also the vulnerabilities of the services that provide AIS data. Many users in the maritime community receive and process live AIS data via web-based

portals. Actors with the necessary resources could cyberattack those web-based portals and block or corrupt such information.

D. MERGING AND CLUSTERING OWNERSHIP DATA

Vessel IMO numbers provided by AIS include several pieces of company-related information: company name, company IMO number, company address, and company phone number. There can be inconsistencies in that data such as typographical errors, spelling errors, and differences in capitalization and punctuation. To check for these in our data sample, clustering methods were used from the open-source software *OpenRefine*. Techniques for clustering included the key-collision methods Fingerprint, N-Gram, Methaphone3, and Cologne-phonetic. Nearest-neighbor methods were also explored including the edit distance (e.g., the Levenshtein distance) and prediction by partial matching. More information about the clustering technique in *OpenRefine* can be found on the “Git” page [22]. We applied these clustering techniques to the company names, addresses, and phone numbers.

Unfortunately, finding useful suggestions via such methods was rare. For company names, the methods would often match common words among maritime companies such as shipping, maritime, marine, management, ltd, co, bulk, transport, group, carrier, Asia, and Pacific. Table 2 is one such example. The two company names only share generic terms “Ship,” “Management,” “Pte,” and “Ltd.” Therefore, they should be considered unaffiliated.

Table 2. Generic Terms Example

ISM Ship Management Pte Ltd	5576575	03-03 Vertex,33, Ubi Avenue 3,Singapore 408868	+65 6222 9077
Thome Ship Management Pte Ltd	1185781	43-01,Hong Leong Building,16, Raffles Quay, Singapore 048581	+65 6220 7291

Using the aforementioned clustering techniques on company addresses resulted in false positives exclusively. Suggested clusters included addresses that were different only in postal code as in Table 3, or addresses located on different floors of the same building

as in Table 4. In Table 3, two companies have addresses which are only distinguishable by postal code. However, there is a lack of other evidence to suggest affiliation. In Table 4 two companies are located in the same building but on different floors. There was no other evidence to suggest affiliation, so the two companies were considered unaffiliated.

Table 3. Different Postal Code Example

Iolcos Hellenic Maritime Enterprises Co Ltd	1077535	Kifisia,145 63Athens, Greece	+30 210 623 3960
Consolidated Marine Management Inc (CMM)	1754044	Kifisia,145 62Athens, Greece	+30 210 459 5100

Table 4. Different Floor Example

2GO Group Inc	415289	15th Floor, Times Plaza Building,Ermita,1000Manila, Philippines.	+632554 8777
Magsaysay Shipmanagement Inc	5304442	20th Floor, Times Plaza Building, United Nations Avenue,1000Manila, Philippines.	+632526 8888

While similarities like those in Tables 3 and 4 can suggest affiliation, in the absence of other evidence they are inconclusive. It is not uncommon to find companies in the same industry physically located nearby. Business parks and multiuse buildings often congregate businesses of similar types, but this does not mean that those businesses are affiliated.

Also, companies sometimes shared the same incomplete address. Table 5 is an example where the addresses were both missing a building or unit number to distinguish them. Without such information, and with no other evidence in company name or phone number, those cases were also considered as unaffiliated.

Table 5. Missing Building or Unit Number Example

White Panama SA	946350	Namikata-cho,Imabari-shi, 799-2101,Japan.	+81 898 417830
Yano Kaiun Co Ltd	528783	Namikata-cho,Imabari-shi, 799-2101,Japan.	+81 898 418188

One of the most effective factors to cluster companies is keywords in the company name. Excluding generic terms as mentioned, some companies had only minor differences in name. Table 6 is one such example; the address, phone number, and company IMO number are all different but the company name is the same except for “(HK),” an abbreviation for Hong Kong. It is common for larger companies to have multiple locations and this would explain their different addresses and phone numbers. The likelihood that different companies in the same industry in the same region of the world have the same name is small.

Table 6. Affiliates Example

Misuga Kaiun (HK) Ltd	5157983	Room 2601–02, Island Place Tower, North Point, Hong Kong	+852 3420 2330
Misuga Kaiun Co Ltd	1773755	1692-2, Nakanosho-cho, Shikokuchuo-shi, 799–0422, Japan.	+81 3 3261 6725

There were also a few company pairs with the same phone number, albeit with differences in the spacing between digits, while the addresses, company names and company IMO numbers were different. In such cases the companies were merged. In some cases, companies shared information across more than one attribute, such as keywords in the company name and locations within the same region. The similar names give greater justification for merging. Table 7 is one such example.

Table 7. Affiliates Example

Carras (Hellas) SA	606585	65, Akti Miaouli, 185 36 Piraeus, Greece.	+30 210 458 7000
Chandris (Hellas) Inc	1248931	95, Akti Miaouli, 185 38 Piraeus, Greece.	+30 210 458 4000

In many cases, the appearance of narrow keywords among company names was deemed sufficient to indicate association. Keywords shared among companies in the same industry could indicate that they are subsidiaries of a larger parent or conglomerate.

For example, Table 8 shows companies that share “V Ships” in their name. This term was deemed as non-generic and a useful keyword within the company names. All such pairs found were verified to belong to parent companies. However, many more subsidiaries have dissimilar names. The complete holdings of conglomerate maritime companies of interest are not published. Many of the companies observed are private and their holdings are not disclosed publicly.

Table 8. Affiliates Example

V Ships France SAS	1971011	34, place Viarme, 44000 Nantes, France.	+33 2 28 09 34 20
V Ships Ship Management (India) Pvt Ltd	5433068	Unit S005, Ground Floor, Delta Wing, Raheja Towers, 177, Annasalai, Chennai, 600002, India.	+91 44 4293 4022

E. REPRESENTING CO-OCCURRENCES

Co-occurrence can be represented in several ways. For this research, a co-occurrence is interpreted as two or more vessels in nearly the same place at nearly the same time. One could alternatively interpret it as appearing at the same location at very different times; this would make sense if a location is highly isolated and two or more entities appearing there is cause for suggesting affiliation. Two occurrences in a narrow time period in a very isolated area, even if large, could also suggest affiliation.

In the maritime domain, large vessels do not appear close together in the same way that people or animals do. The location does affect a vessel’s distance from others; high-density areas like maritime chokepoints and ports force a closer proximity. Taking everything into account, this work set the spatial and temporal thresholds for characterizing co-occurrence events at 7.79 km and one day, respectively. 7.79 km translates to approximately 0.05 decimal degrees. Figure 7 depicts the distance. In the center is the island of Itu Aba, roughly the size of a small runway. Figure 8 provides a closer view to give a better sense of the scale.



Figure 7. Itu Aba Island – 7.79 Km Radius. Adapted from [23].



Figure 8. Itu Aba Island (Zoomed In). Adapted from [24].

The Haversine formula was used to compare two locations (in latitude and longitude), producing a great-circle difference between the two points in kilometers. The formula allows us to correct for the decrease in distance between longitude degrees with latitude, and also correct for the curvature of the earth.

Another co-occurrence factor is how many different co-occurrence events are shared between pairs. One way to define a different co-occurrence is if the distance is more than our spatial threshold (7.79 km). Counting different co-occurrences allows for more fine-grained analysis about co-occurrences among vessels. As previously mentioned in Chapter II when discussing [2], [10] and [11], co-occurrences among pairs in more than one place indicate a greater likelihood of affiliation.

We also calculated the total number of co-occurrences between pairs of vessels. Like the count of different locations, this provides an additional means to analyze the relationship between them. We calculate both the number of different locations in which they appear together and the total number of times they appeared together.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. RESULTS

For random vessels, selected locations of interest, and selected vessels of interest, co-occurrences were identified among the vessels. Afterward, vessel-ownership data was acquired and merged with the results. If the pairs' owners were the same, the pair was assigned a co-occurrence value of one, zero otherwise.

With all three kinds of data, positive confirmation of affiliation by co-occurrence was rare. Many more co-occurring pairs did not share the same owner as pairs that did. The random pairs had only 1% with the same owner; the Hong Kong locations-of-interest pairs had only 8%; and the Spratly locations-of-interest pairs had only 9%. The COSCO vessels-of-interest data showed a much higher percentage, 71%, but that was to be expected; if all companies in the COSCO group were assumed to be subsidiaries, the number would be 100%. In addition, for each of the three kinds of data, most pairs had only a single co-occurrence. That is, most pairs of vessels appeared in only one location and one occasion together.

A. EXPERIMENT I: RANDOMLY SELECTED VESSELS

We wanted to first see if there was a correlation between the number of places a pair of vessels reported together and the number of total times the pair had reported together. In Figure 9, all five runs finding random vessels were compiled together, and a scatter plot was produced. The x-axis shows the number of unique locations a pair of vessels reported together and the y-axis shows the total number of times a pair of vessels reported together. The Pearson correlation coefficient score was 0.68.

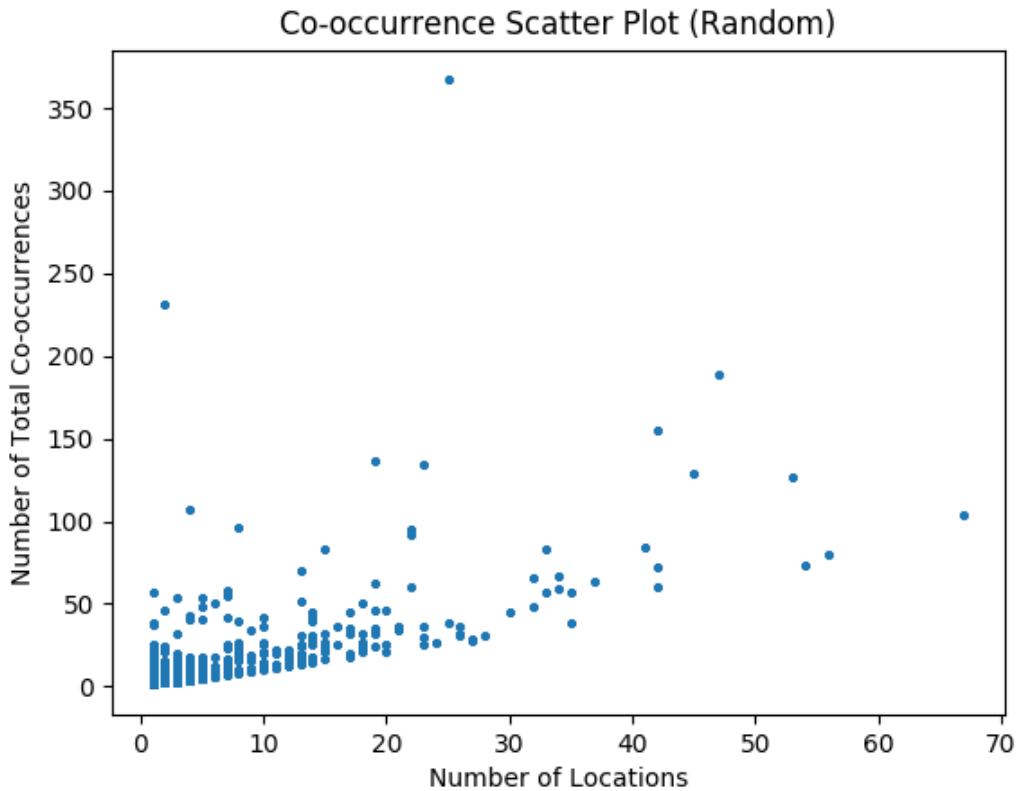


Figure 9. Co-occurrence Scatter Plot (Random Vessels) ($r = 0.68$)

This indicates a weak correlation between the two variables. The bottom of Figure 9 shows a well-defined floor of the plotted points as a result of numbers being counted twice. If a pair of vessels reported co-occurrences in ten places, then they must have reported at least ten co-occurrences. So as the number of unique locations goes up, the total number of co-occurrences must at least go up by as much. The many points at that minimum say that pairs of vessels that co-occur tend to do so at different locations and times.

To give a sense of the results for the random iterations, Figures 10 and 11 present histograms of the locations and the number of co-occurrences. In the x-axis of each histogram is either the number of locations a pair of vessels reported together or the number of times a pair of vessels reported together. The y-axis of each histogram is the normalized count.

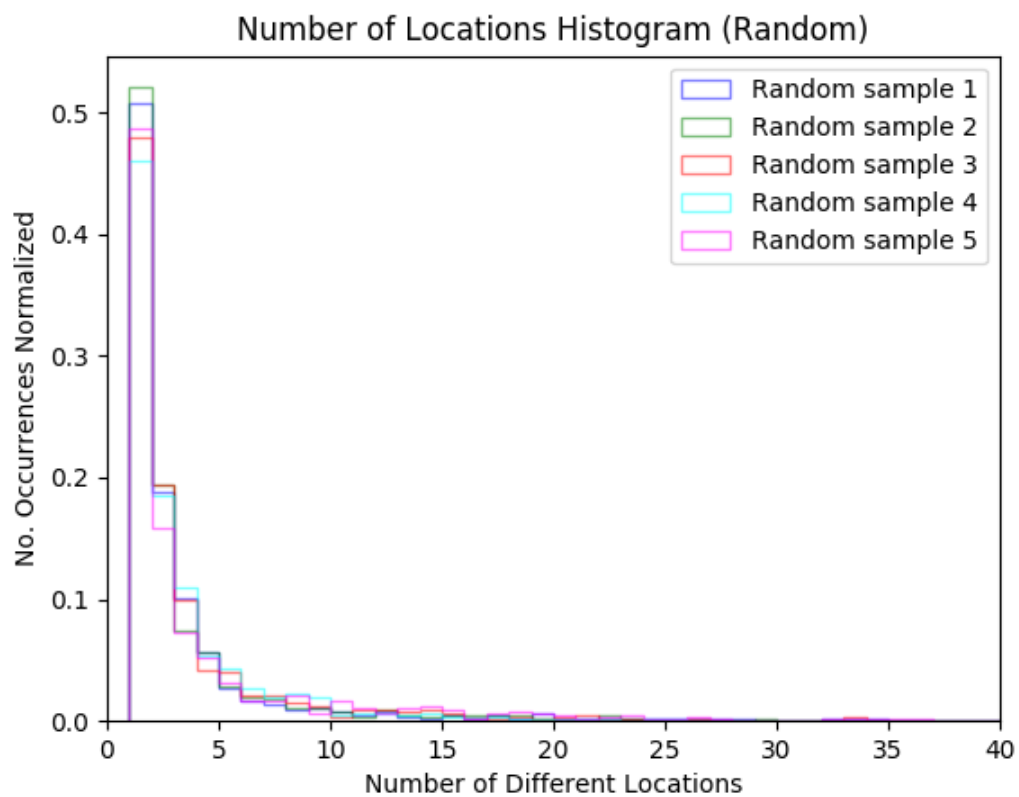


Figure 10. Number of Locations Histogram (Random)

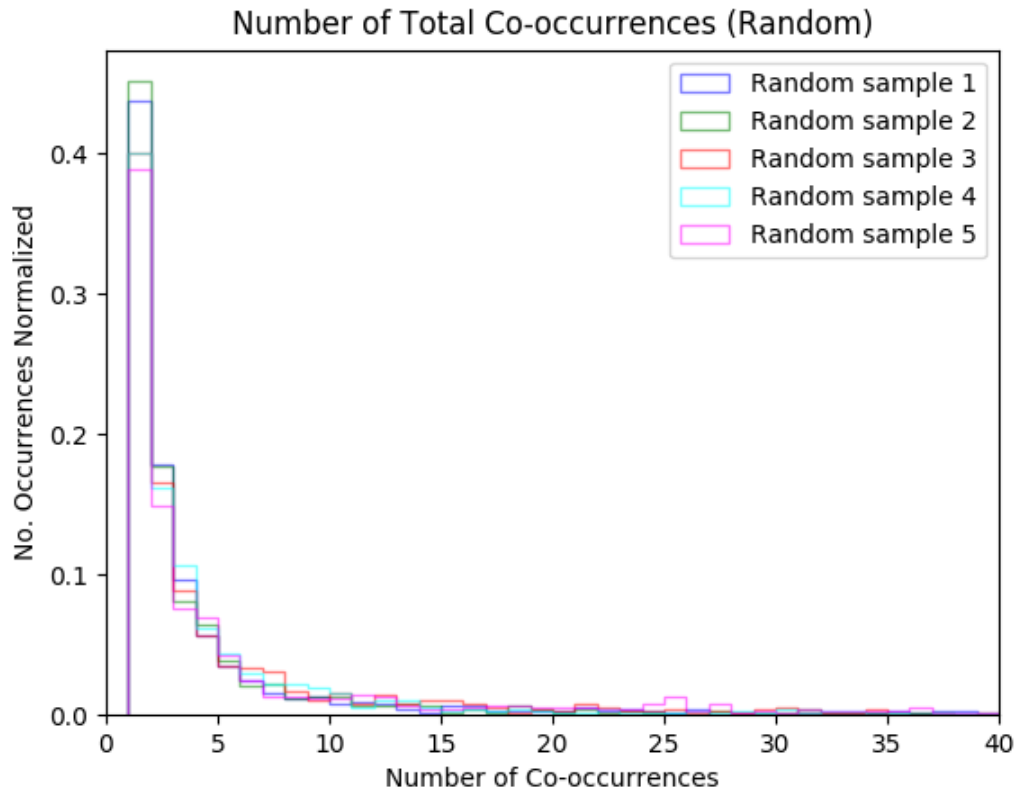


Figure 11. Total Co-occurrences Histogram (Random)

There appears to be much similarity between the five runs in the histograms. As previously mentioned, most pairs have only one co-occurrence event. In Figure 10, all five iterations show approximately 50% of the pairs to have only one co-occurrence. In Figure 11, the number of pairs with one total co-occurrence is between 40–50%. The numbers drop off significantly after that, with 20% or less having two co-occurrences.

To see if greater co-occurrence indicates greater likelihood of affiliation, we plotted a Receiver Operating Characteristic (ROC) curves. This measures the true positive rate against the false positive rate. The true positive rate is defined as the percentage of positives correctly identified, and the false positive rate is the percentage of negatives correctly identified. In this research, a true positive is a pair of vessels that meet or exceed a threshold of co-occurrences, and the pair are confirmed to be affiliated. For example, if the threshold is five total co-occurrence events and a pair has seven, then the

pair exceeds the threshold; if the pair is confirmed to be affiliated via the company data, then we consider the pair a true positive. We do this for locations (Figure 12) and number of co-occurrences (Figure 13). For each of total co-occurrences and unique locations, we vary the threshold from the min to max of each sample; this allows us to produce the curves that appear in the ROC graphics. The upward bulge of the curves says that the number of co-occurrences a pair of vessels has is at least better than random at predicting affiliation.

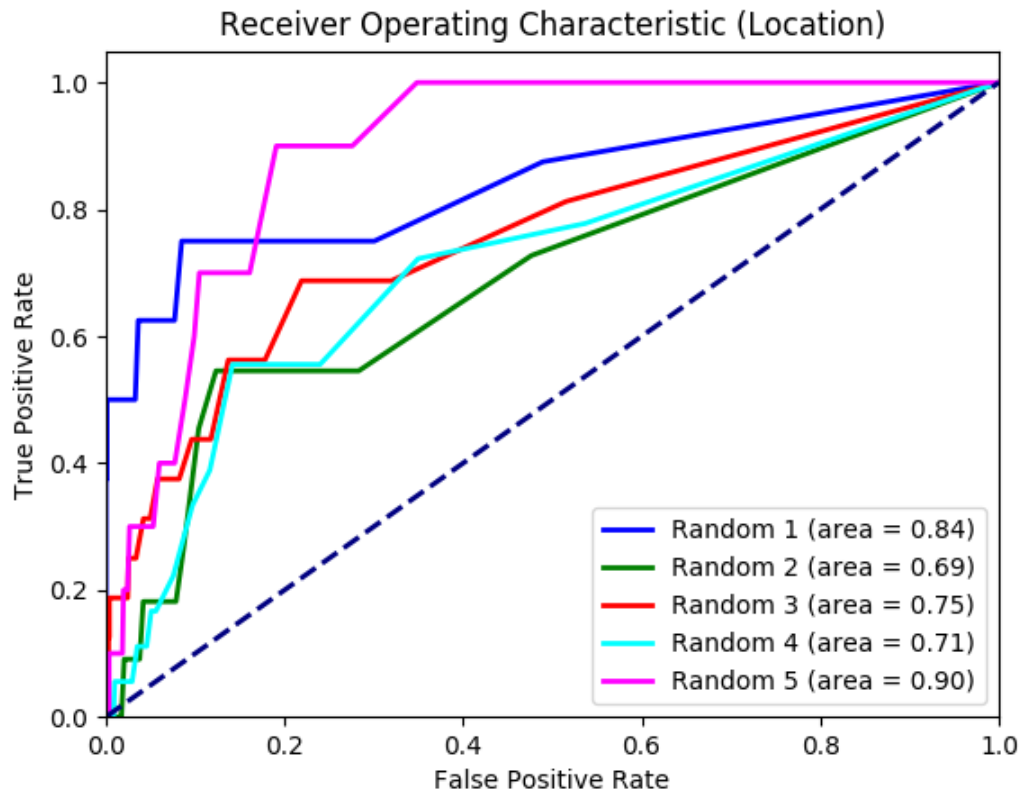


Figure 12. ROC Unique Locations (Random)

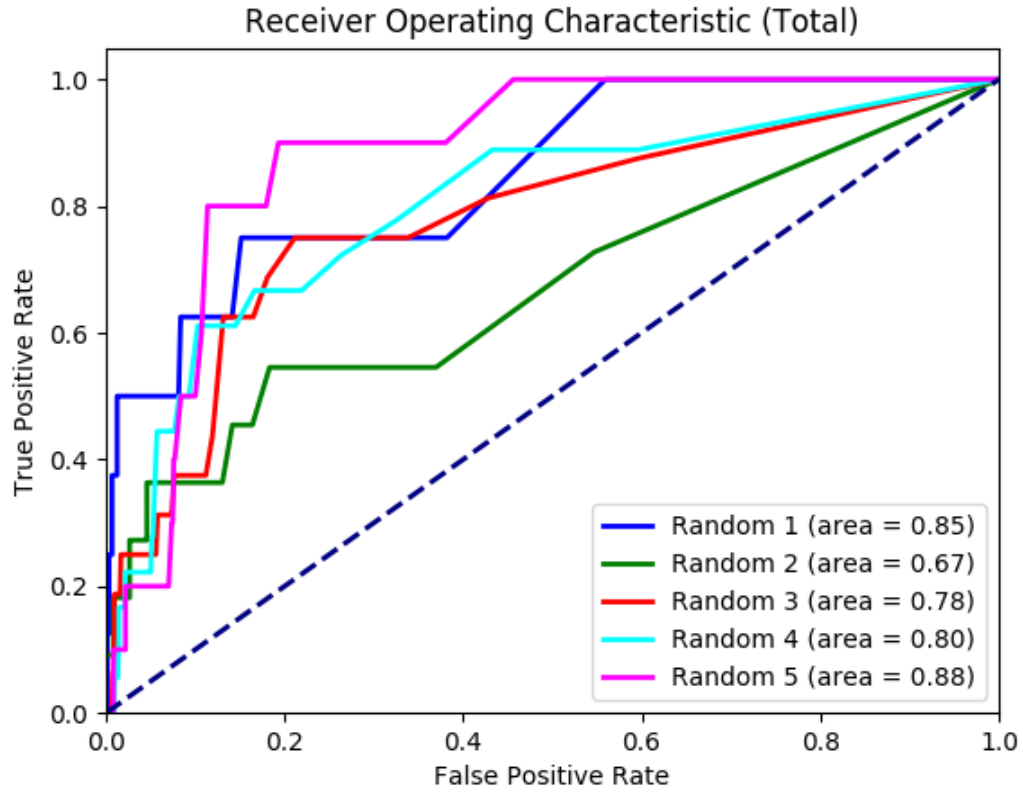


Figure 13. ROC Total Co-occurrences (Random)

B. EXPERIMENT II: LOCATIONS OF INTEREST (LOI)

Figure 14 presents the correlation between number of locations and number of co-occurrences for the Hong Kong sample, and Figure 15 for the Spratly Islands sample. The Pearson correlation coefficient was 0.8 for the Hong Kong data and 0.07 for the Spratly Island data.

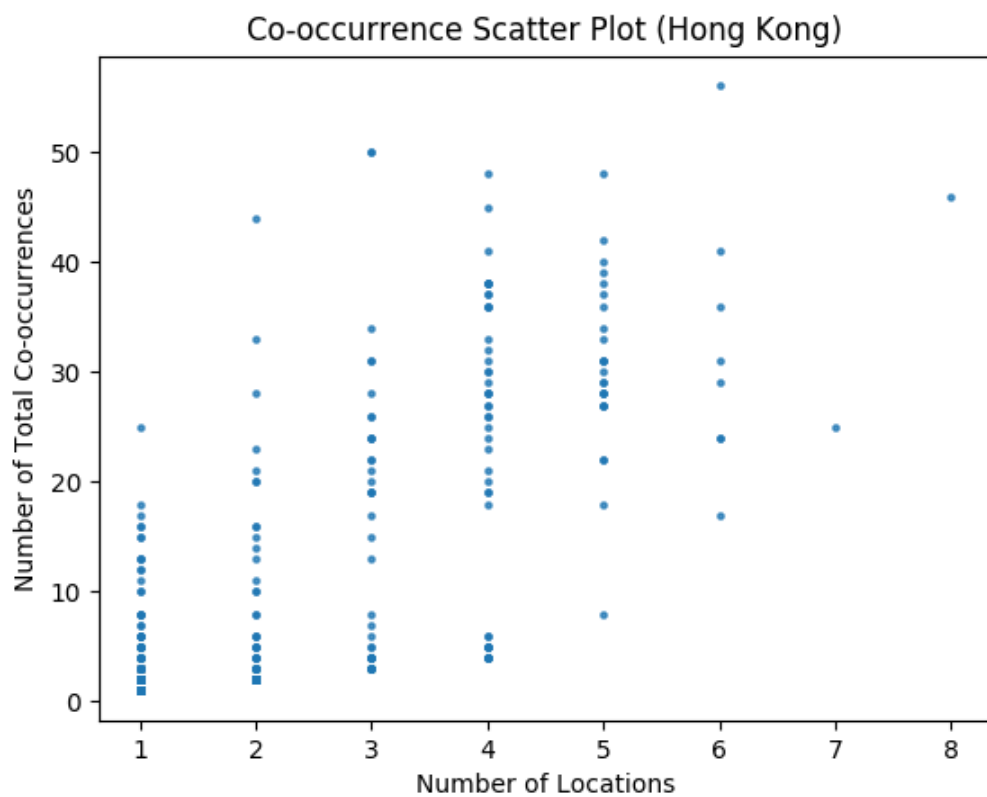


Figure 14. Co-occurrence Scatter Plot (Hong Kong) ($r = 0.8$)

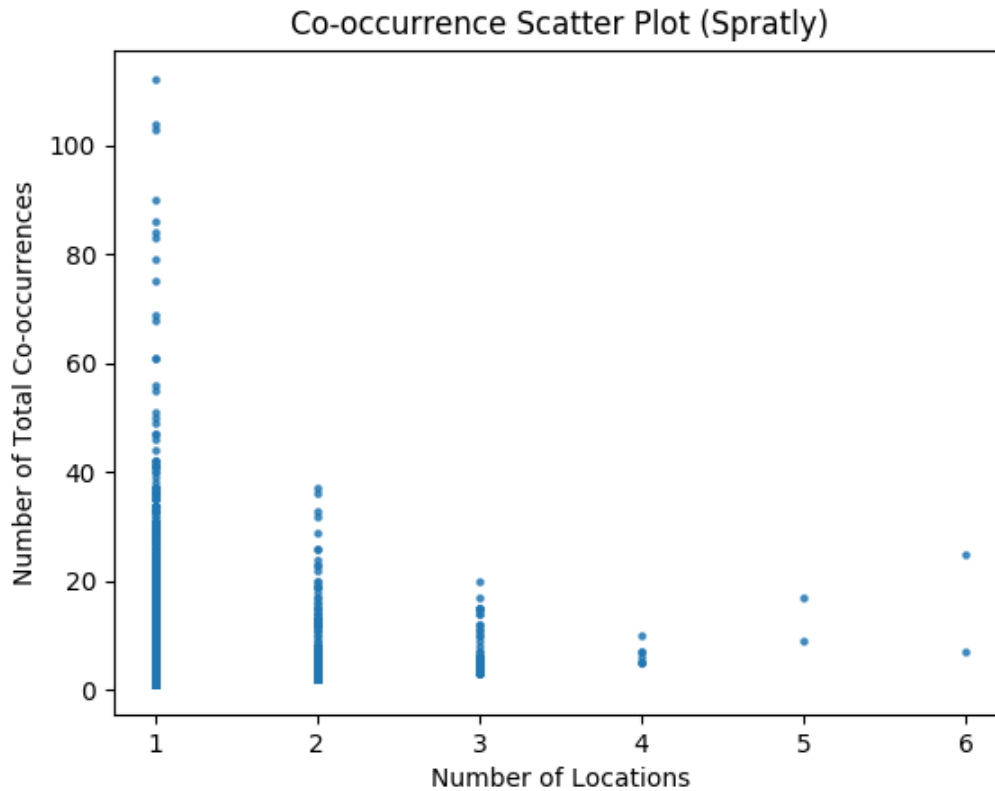


Figure 15. Co-occurrence Scatter Plot (Spratly) ($r = 0.07$)

Note that near Hong Kong the number of locations is strongly related to the number of co-occurrences, but not near the Spratly Islands. The Hong Kong data sample had 213 different vessels whereas the Spratly Islands data sample had 65, but it seems unlikely that this affected the results. We conclude that the demographics and behaviors of vessels in these areas are different. In Figure 15 we can see that a few of the pairs appeared together in only one place on over 100 occasions, which causes the low correlation score. One would expect a pair of vessels that only have a single unique location to have a much lower total number of co-occurrences. Because these small islands are isolated, perhaps the vessels have fewer places to go. In other words, they have greater reason to operate in a smaller area, or there are fewer options for mooring/anchoring. These are young islands being enhanced in various ways, so perhaps the vessels are working together in a small area over long periods of time. For example,

perhaps there are two or more vessels coordinating a dredging operation. While it is uncertain what the causal factors are for the difference in correlation, we can observe a somewhat different position reporting behavior among vessels in different locations.

Figures 16–19 are histograms of the values for locations and number of co-occurrences for pairs of vessels in the Hong Kong and Spratly Islands data samples. As we saw using randomly selected vessels, most pairs (80-90%) only co-occurred in one location.

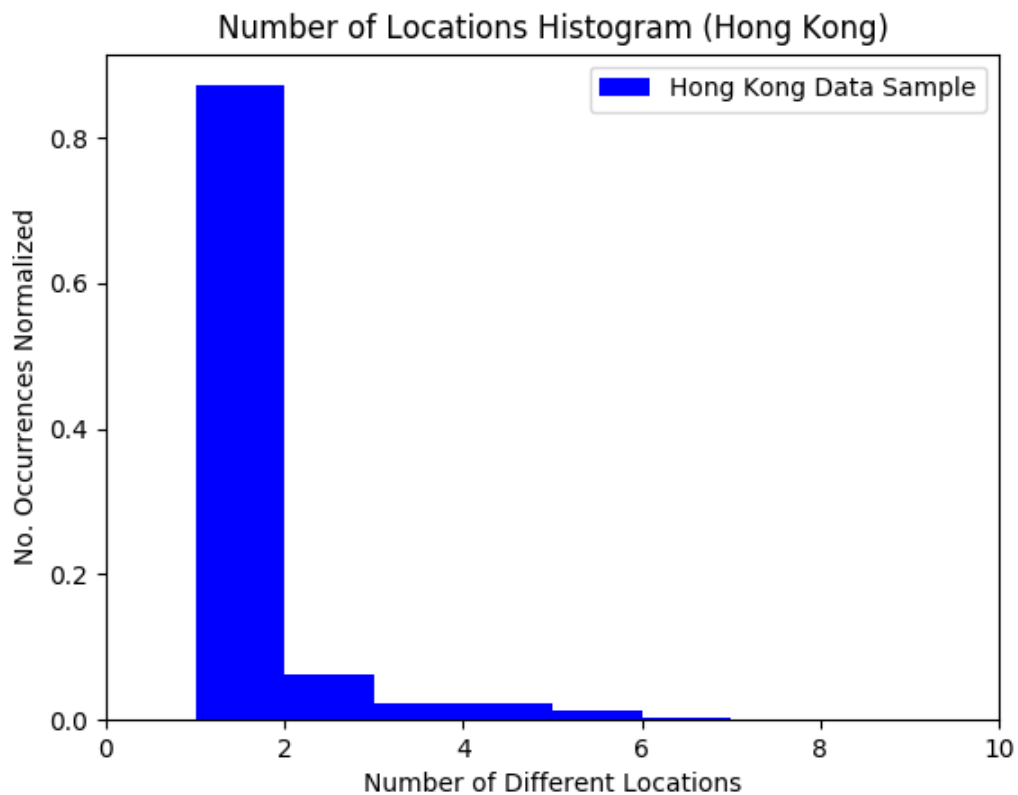


Figure 16. Unique Locations Histogram (Hong Kong)

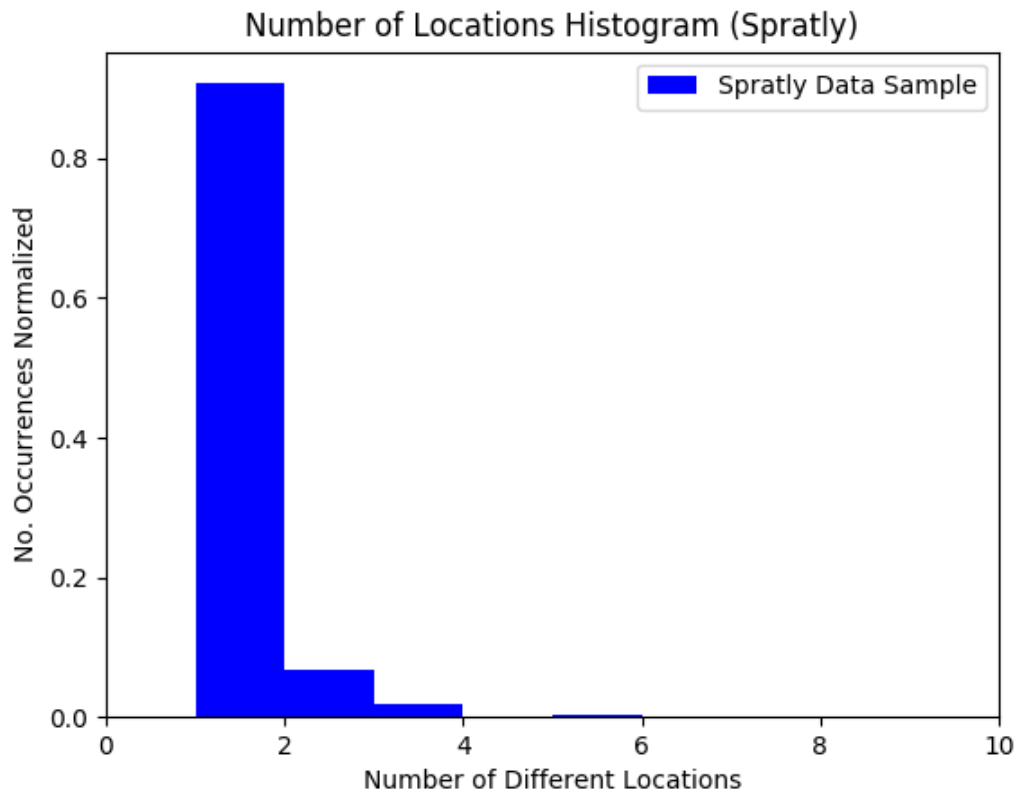


Figure 17. Unique Locations Histogram (Spratly)

Figures 18 and 19 compare the two data samples. In the Hong Kong data, nearly 80% of the pairs had only one co-occurrence; in the Spratly Islands data, it was 35%.

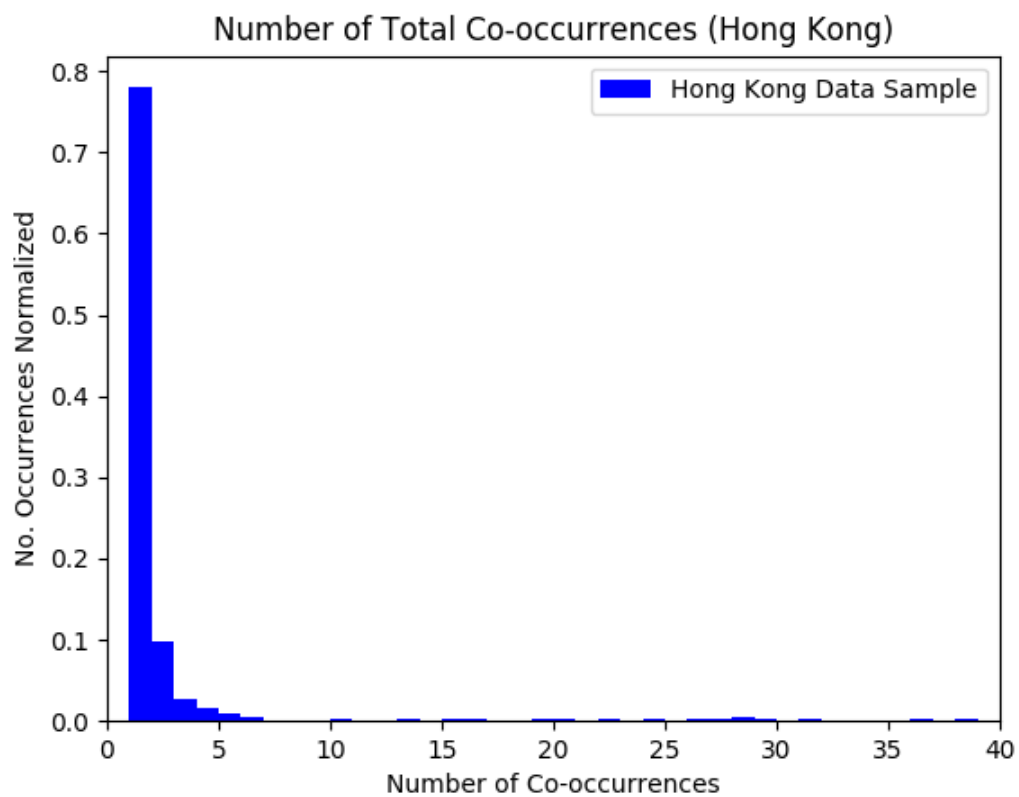


Figure 18. Total Co-occurrences Histogram (Hong Kong)

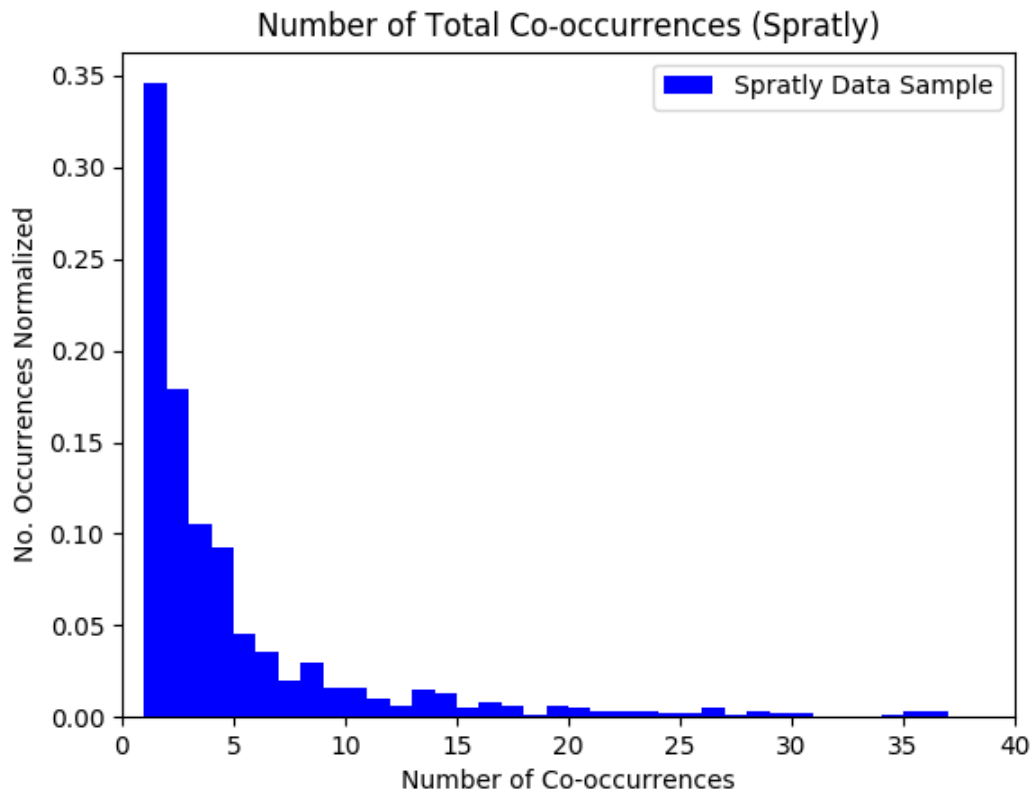


Figure 19. Total Co-occurrences Histogram (Spratly)

Figures 20–23 show the ROC curves for unique locations and total number of co-occurrences for each of the two data samples. In the Hong Kong sample, the area under the curve is 0.78; in the Spratly Islands sample, the area is only 0.52. The Hong Kong results were more consistent with the random samples.

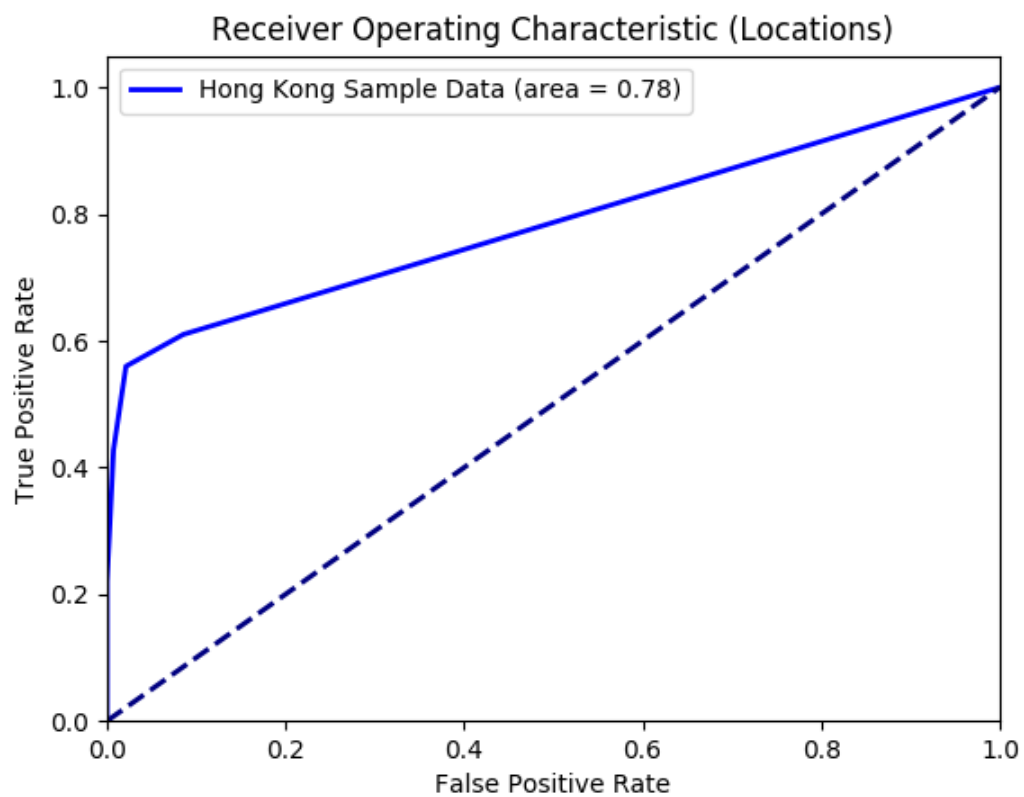


Figure 20. Unique Locations ROC (Hong Kong) (AUC = 0.78)

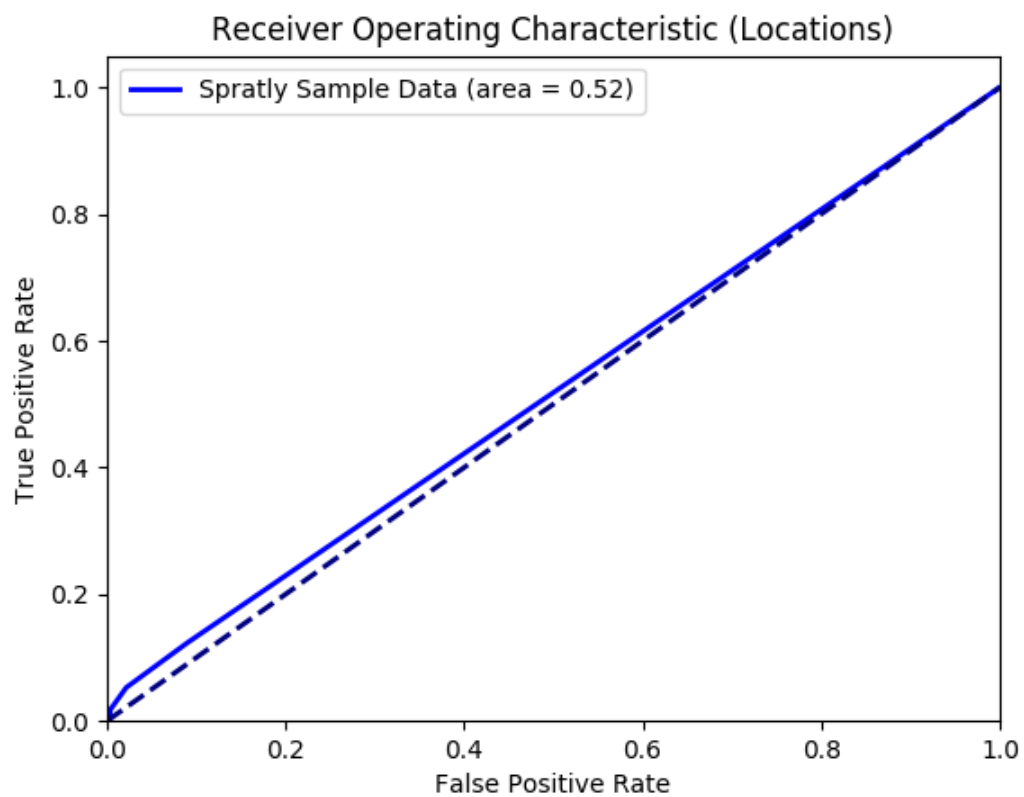


Figure 21. Unique Locations ROC (Spratly) (AUC = 0.52)

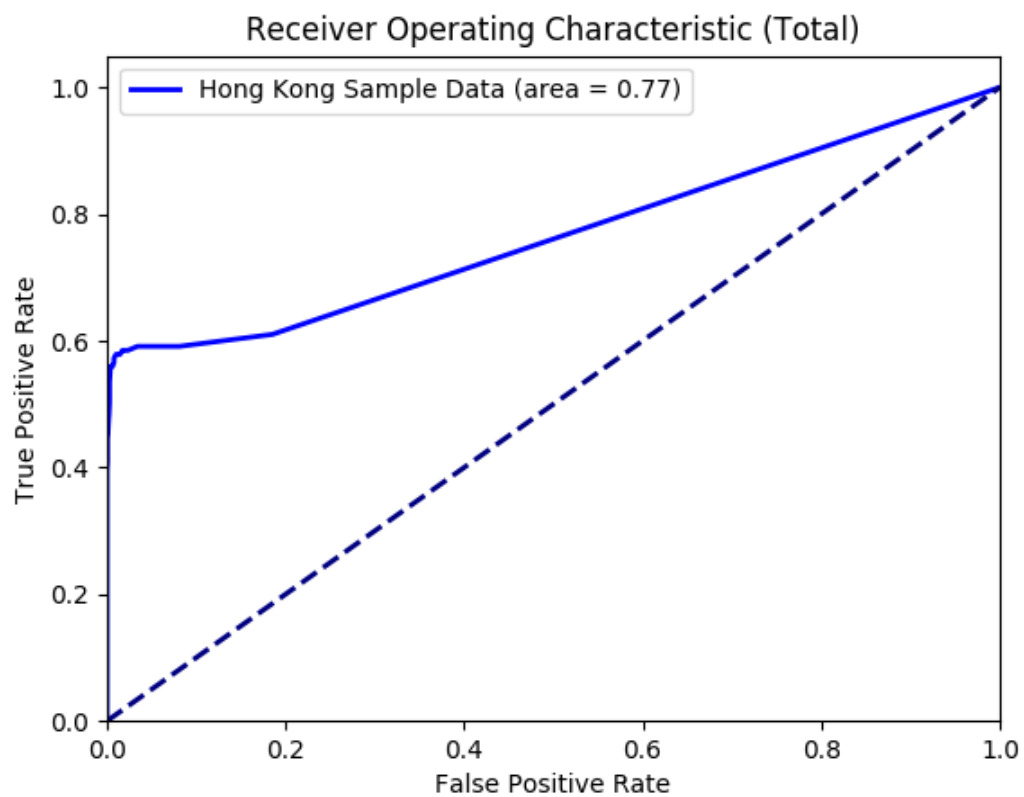


Figure 22. Total Co-occurrences ROC (Hong Kong) (AUC = 0.77)

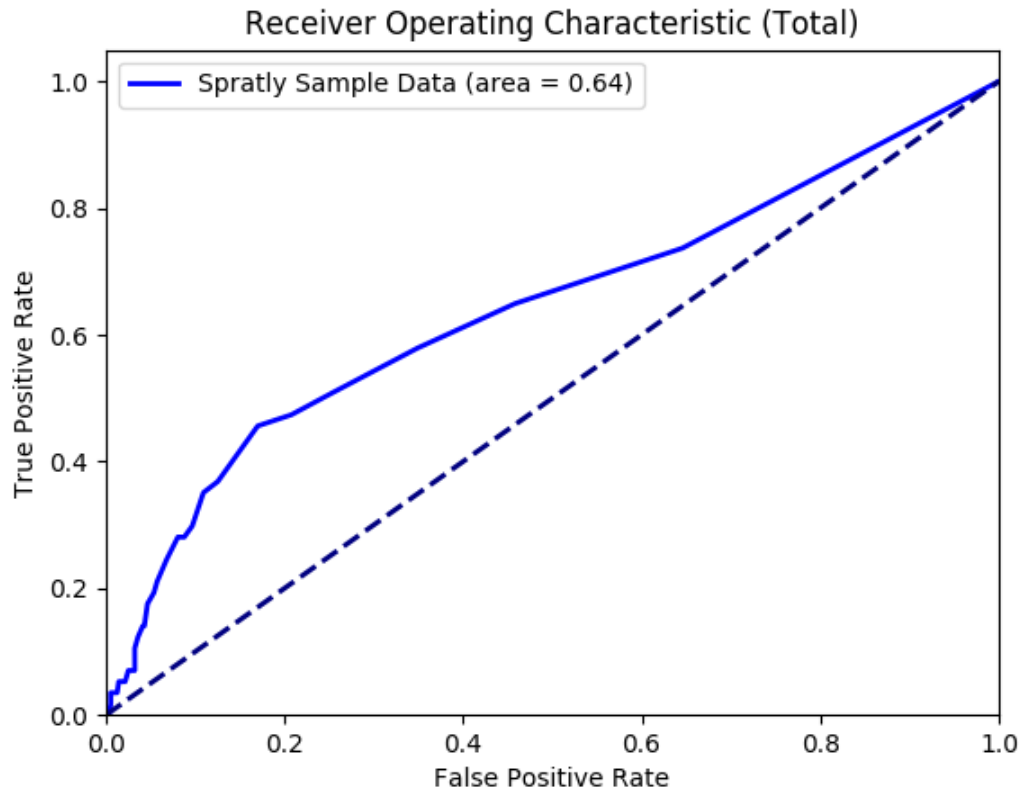


Figure 23. Total Co-occurrences ROC (Spratly) (AUC = 0.64)

While sample sizes are small, it appears that there is some difference in the reporting behavior of vessels at different locations. In densely populated areas like the port of Hong Kong, the results are more in line with the random data samples. The higher number of co-occurrence events provides a higher level of confidence that a pair of vessels are affiliated. The Spratly Islands sample does not provide much confidence at all; data suggests only slightly better than random chance that a pair of vessels is affiliated. This seems counterintuitive and contradictory of other work on co-occurrence. The Spratly Islands are more isolated and higher numbers of co-occurrence in isolated areas would suggest a better chance of affiliation than high numbers of co-occurrence in areas that are densely populated.

C. EXPERIMENT III: VESSELS OF INTEREST (VOI)

We can also start with vessels of interest. In our experiments, we started with a maritime conglomerate company COSCO specializing in bulk transportation. The vessels of COSCO and its affiliate companies were used as the vessels of interest in a known maritime network.

Figure 24 is a scatter plot of the COSCO data sample. The Pearson correlation coefficient was high at 0.9 between the number of locations a pair of vessels report and the number of times a pair of vessels reports. There were 298 vessels in the COSCO data sample. The reporting behavior was relatively predictable; vessels that report a co-occurrence rarely report more than one co-occurrence event at a given location. This was a much stronger trend than in the other data samples.

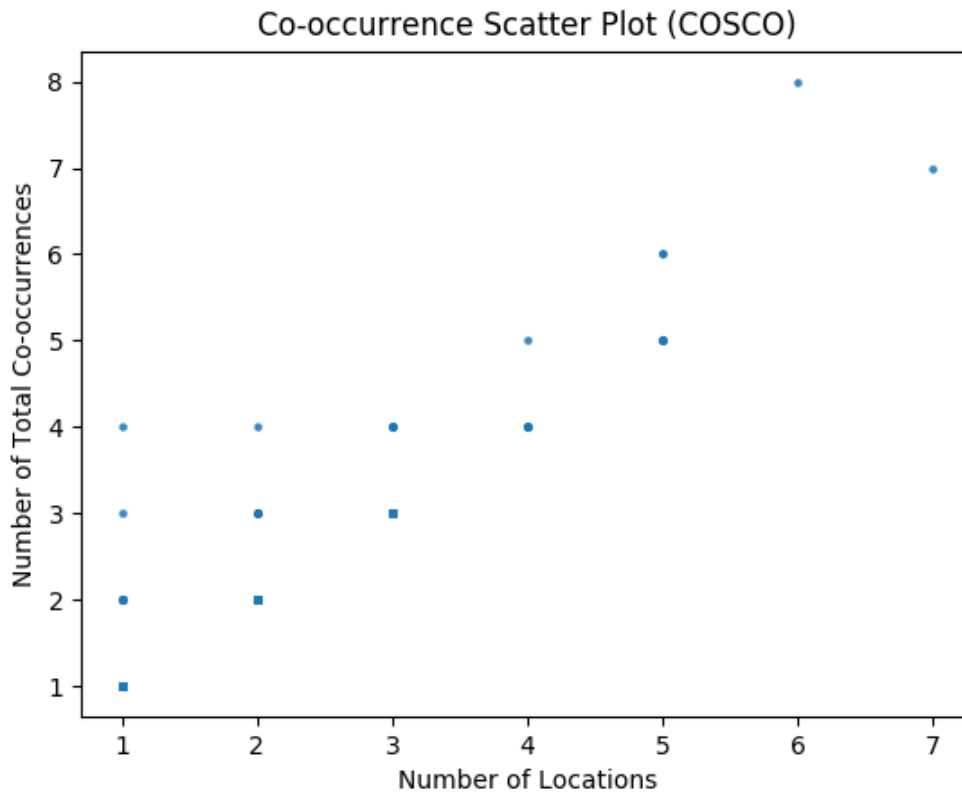


Figure 24. Co-occurrence Scatter Plot (COSCO) ($r = 0.9$)

Figures 25 and 26 show a similar trend. These histograms show that nearly 80% of the COSCO sample reports a co-occurrence in only one location and only on one occasion. Most of the sample is transport vessels, and these vessels do not have much reason to congregate or to loiter in an area like those in the Spratly sample. Instead, the COSCO vessels appear to be intent on getting from one destination to the next as quickly as possible.

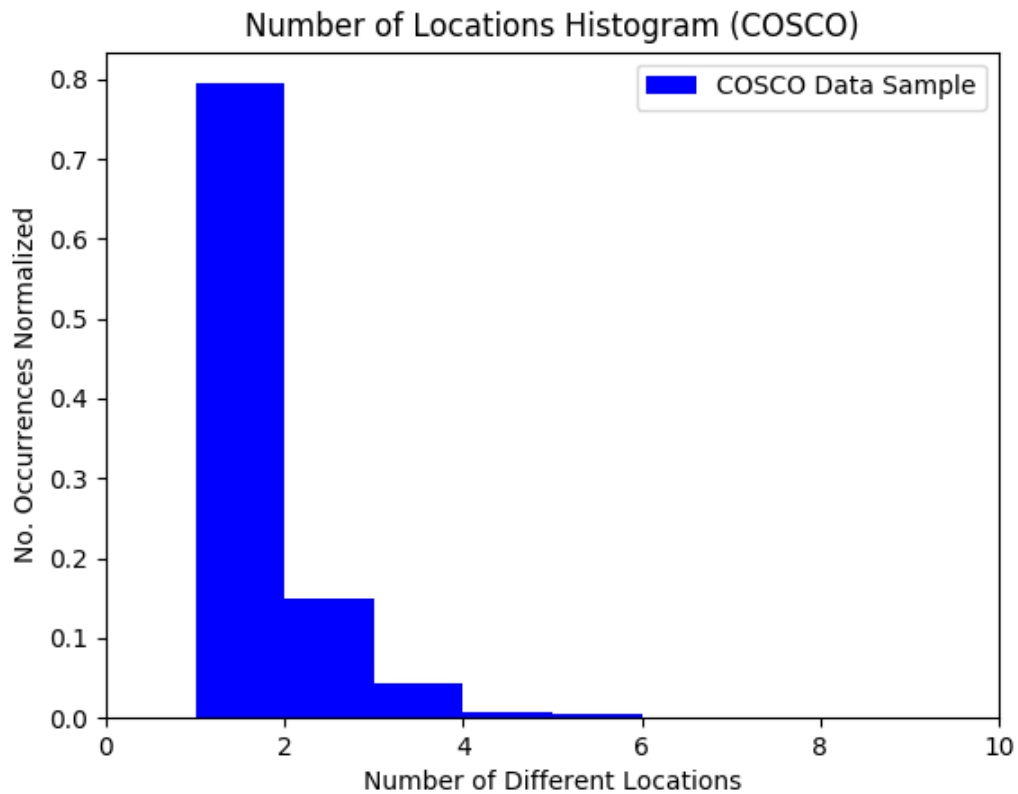


Figure 25. Unique Locations Histogram (COSCO)

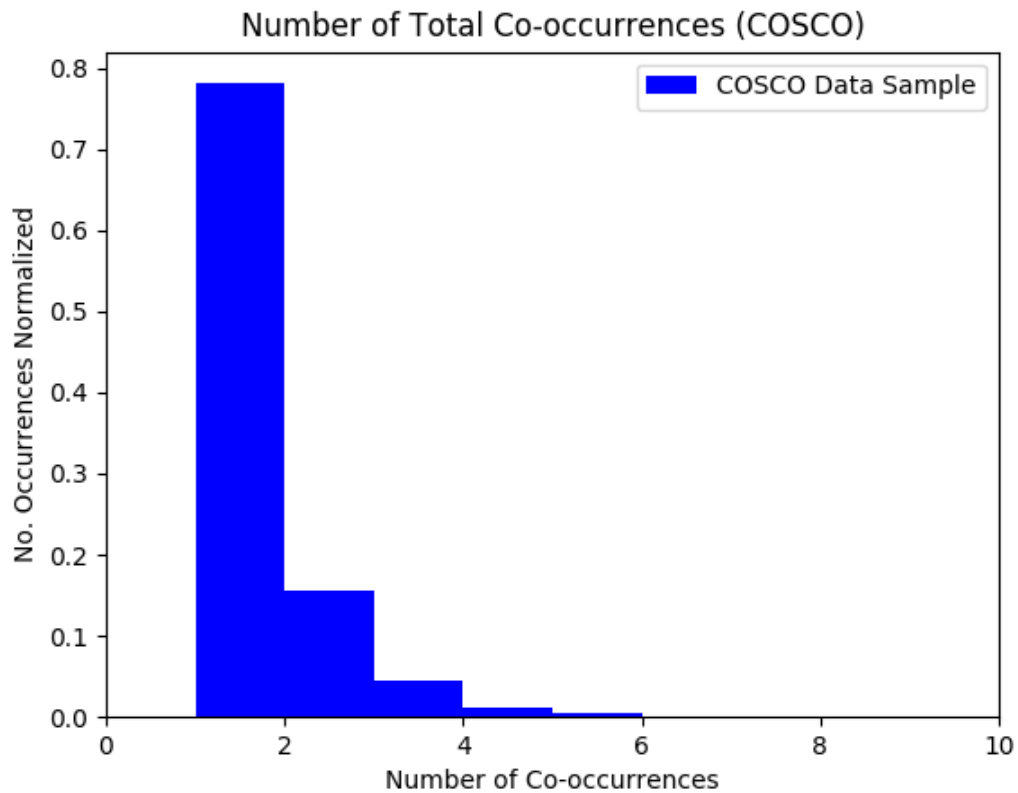


Figure 26. Total Co-occurrences Histogram (COSCO)

Figures 27 and 28 are the ROC curves for locations and co-occurrences for the vessels-of-interest data. Both curves have low areas under the curve, at 0.51. This suggests, like the Spratly Islands sample, that the predictive power of each variable is no better than randomly guessing that a pair of vessels is affiliated.

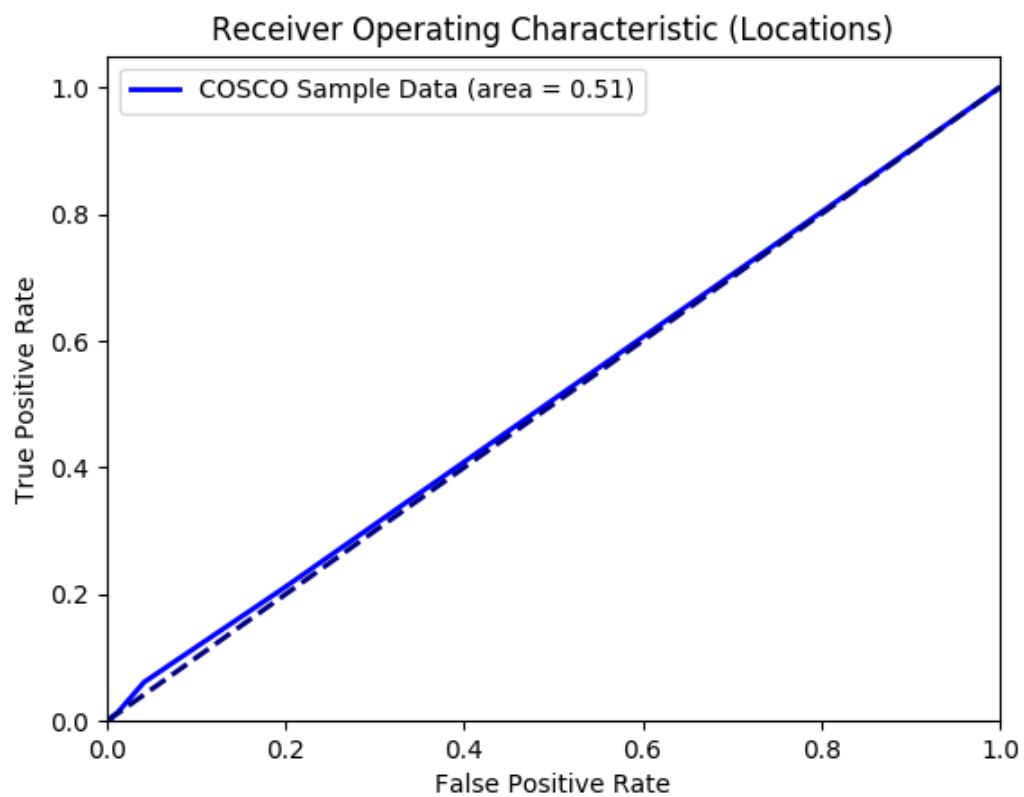


Figure 27. Unique Locations ROC (COSCO) (AUC = 0.51)

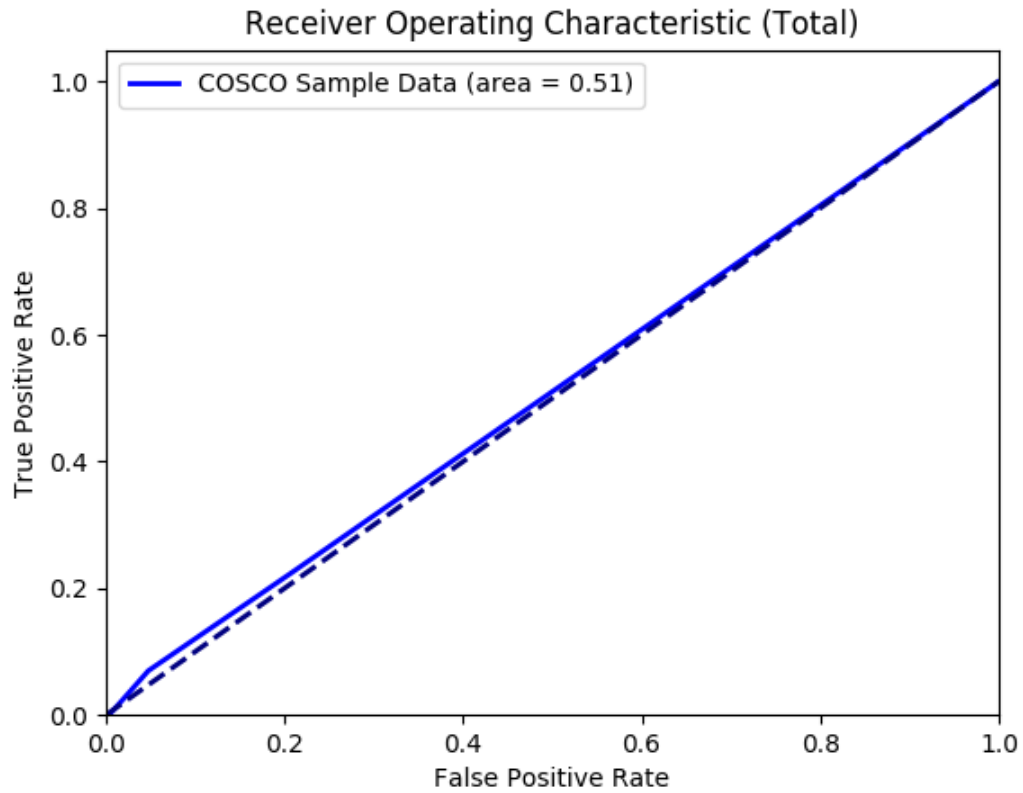


Figure 28. Total Co-occurrences ROC (COSCO) (AUC = 0.51)

Figures 29 and 30 present the combined ROC results for each of the samples. We can see that there are different trends of co-occurrence. Some vessels report together in one place many times, while other vessels report together in one place on one occasion.

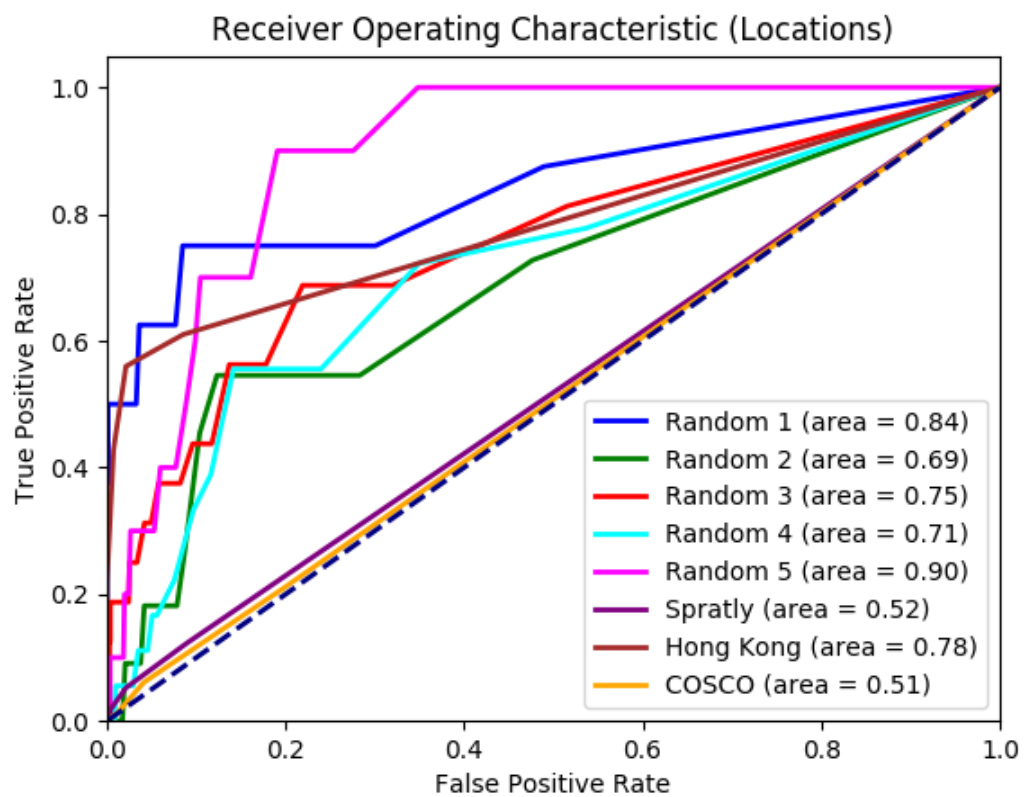


Figure 29. Unique Locations ROC (All Samples)

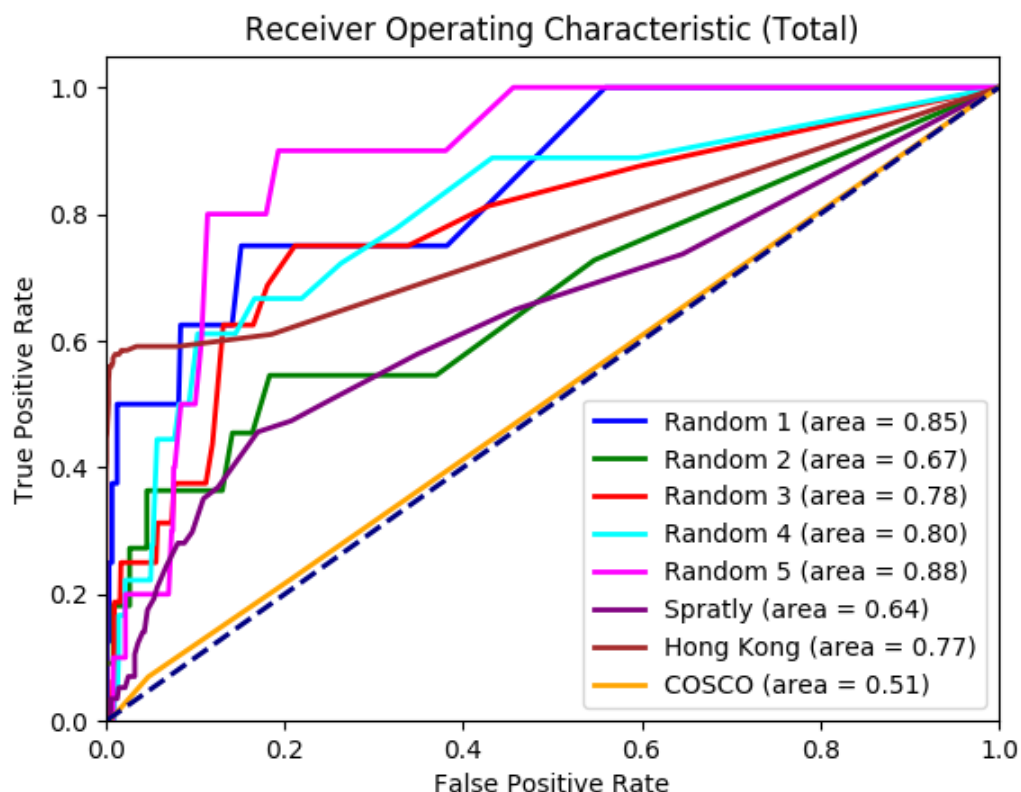


Figure 30. Total Co-occurrences ROC (All Samples)

D. NOTABLE ASSOCIATIONS

There are potential insights to be gained by observing individual pairs of co-occurrence events. What follows are observations on several of the pairs in each of the samples.

Table 9 shows three pairs of vessels extracted from the compiled results for the five iterations of randomly selected vessels. The random samples generally have higher amounts of co-occurrence events, both in unique locations and total number. It is not apparent why; it could be due to chance, or it could be some causal factors. Perhaps certain vessels transmit their position with higher frequency, and random selections are more likely to grab these vessels. Another observation is the erroneous MMSI number “11111111.” As previously mentioned, AIS has several faults, and the ability of vessels to transmit false MMSI numbers deliberately or mistakenly is one such fault.

Table 9. Interesting Pairs (Random Samples)

MMSI 1	MMSI 2	Locations	Total	Affiliated	Company 1	Company 2
111111111	548438100	25	367	0	LD Ports & Logistics Pte Ltd	Cokaliong Shipping Lines Inc
548042500	548025500	53	127	1	Solid Shipping Lines Corp	Solid Shipping Lines Corp
548856000	548293100	67	104	0	Transnational Uyeno Maritime Inc	Magsaysay Shipmanagement Inc

Table 10 shows three pairs of vessels from the Hong Kong sample. The sample includes the pair that had the highest amount of unique locations of eight, and the pair that had the highest number of total co-occurrences of 56. The company “Chu Kong High-Speed Ferry Co Ltd” appears in many of the pairs; it is not surprising that ferry company vessels have the most co-occurrences in a densely populated area like Hong Kong.

Table 10. Interesting Pairs (Hong Kong)

MMSI 1	MMSI 2	Locations	Total	Affiliated	Company 1	Company 2
477197300	477197200	6	56	1	Chu Kong High-Speed Ferry Co Ltd	Chu Kong High-Speed Ferry Co Ltd
477937500	477937400	8	46	1	Chu Kong High-Speed Ferry Co Ltd	Chu Kong High-Speed Ferry Co Ltd
477937400	477995253	2	23	0	Chu Kong High-Speed Ferry Co Ltd	Callany Ltd

Table 11 shows a few pairs of vessels from the Spratly Islands sample. The top line is the pair of vessels that had the most co-occurrences at 112. It is interesting that this pair had such a high number of total co-occurrences, but in only one unique location. This pattern was unique to the Spratly sample as discussed in Chapter V Section B. The pair also did not contain any company information; missing company information occurred in every sample. Outside of the scatter plots, any pairs with missing company data were not included in graphs. Another observation is the high percentage of Chinese vessels in the Spratly sample. Again, this is perhaps not surprising given that Chinese SOE have been observed conducting reef and island enhancement in the area.

Table 11. Interesting Pairs (Spratly)

MMSI 1	MMSI 2	Locations	Total	Affiliated	Company 1	Company 2
600013090	412222222	1	112	missing	missing	missing
413021030	413070000	1	37	0	China Yantai Salvage	Guangzhou Salvage Bureau of the Ministry of Communications PRC
412687000	413100000	6	7	0	China Yantai Salvage	Shenzhen Tonghai Offshore Engineering Co Ltd

Table 12 shows three pairs from the COSCO sample. Included are the pair that had the highest total number of co-occurrences of eight, and the pair with the highest amount of unique locations of seven. What is immediately apparent is the much smaller number of co-occurrences compared to the other samples. As mentioned in Chapter V Section C, transport vessels have little reason to congregate near one another.

The first pair has two companies, which differ in name; however, they were considered affiliated due to the presence of COSCO in their names. The second pair were not considered affiliated, even though a search of “Lianyungang Ocean Shipping Co” reveals that it is a subsidiary of COSCO. The reasons why one pair was considered affiliated while the other was not can be attributed to the process taken in Chapter III Section D. COSCO has many subsidiaries, yet systematically going through each of the pairs is infeasible when dealing with large samples. Having more information on the relationships between companies would provide greater confidence in our ability to infer affiliation.

Table 12. Interesting Pairs (COSCO)

MMSI 1	MMSI 2	Locations	Total	Affiliated	Company 1	Company 2
412342000	412376000	2	2	1	COSCO Bulk Carrier Co Ltd (COSCO BULK)	Shenzhen COSCO LPG Shipping Co Ltd
353801000	412313000	7	7	0	COSCO Bulk Carrier Co Ltd (COSCO BULK)	Lianyungang Ocean Shipping Co
412277000	412148000	6	8	0	Seaspan Ship Management Ltd	Shanghai Ocean Shipping Co Ltd

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSION

A. FINDINGS

The most notable finding has been that co-occurrence among vessels is not manifested uniformly on all occasions. In the COSCO data sample of mostly transport vessels, we saw a different co-occurrence trend than a data sample composed of a random group of vessels. A fleet of bulk carrier vessels likely travels from one port to the next without accompanying vessels, whereas a group of logistics vessels collaborating on a dredging project have reason to cluster in remote areas. The nature and extent of differences in co-occurrence trends requires more study, especially the co-occurrence rates atypical of a category of data.

AIS data may not always be available, and for reasons mentioned in Chapter III Section C, it may not always be accurate either. A lack of data on ownership records or another means to confirm affiliation may compound matters. In situations where AIS is unavailable or untrustworthy, or in situations where there is no confirmation of affiliation, co-occurrence may prove a useful means of suggesting it.

A big obstacle in this research was establishing “ground truth” about vessel affiliation. We attempted to establish it via ownership records but this ignores other ways in which vessels could be affiliated by relationships of managers, operators, and maintainers. Large companies could also have subsidiary companies which could be affiliates, and there may be relationships between stakeholders not explicitly recorded in any documentation. The stakeholders of vessels can also be private companies which are not required to publicly disclose their operations. Having more information about the stakeholders of vessels would provide for a more robust means of establishing a ground truth.

B. FURTHER RESEARCH

One idea worth pursuing is determining the best spatial and temporal thresholds to characterize a co-occurrence event. Different combinations of space and time thresholds might provide other insights about the nature of co-occurrence in the maritime domain.

Another topic worth studying might be exploiting other known maritime networks. It may be worthwhile to study vessels of a different type owned by a company in another industry like fishing. Our preliminary research suggested that co-occurrence trends may be different. If differences are the result of vessel type and the nature of their operations, studying a different form of known maritime network would be helpful.

Lastly, checking co-occurrence among a larger number of vessels would be beneficial. We were limited in taking smaller samples due to the difficulty in acquiring ownership data. Checking co-occurrence between thousands of vessels rather than a few hundred would give more confidence in results.

LIST OF REFERENCES

- [1] United Nations Conference on Trade and Development. “Review of Maritime Transport 2015.” Accessed November 7, 2017. [Online]. Available: <http://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=1374>
- [2] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg, “Inferring social ties from geographic coincidences,” in *Proceedings of the National Academy of Sciences of the United States of America*, 2010. [Online]. doi: 10.1073/pnas.1006155107
- [3] Cyril Ray, Arnaud Grancher, Rémy Thibaud, Laurent Etienne. “Spatio-temporal rule-based analysis of maritime traffic,” in *Conference on Ocean & Coastal Observation*, 2013, pp. 171–178.
- [4] P. Last, C. Bahlke, M. Hering-Bertram, and L. Linsen. “Comprehensive analysis of automatic identification system (AIS) data in regard to vessel movement prediction,” *Journal of Navigation*, vol. 67, no. 5, pp. 791–809, September 2014. [Online]. doi: 10.1017/0373463314000253
- [5] N. Eagle, and A.S. Pentland. “Reality mining: sensing complex social systems.” *Personal and ubiquitous computing*, vol. 10, no. 4, pp. 255-268, May 2006. [Online]. doi: 10.1007/00779-005-0046-3
- [6] N. Eagle, A.S. Pentland, and D. Lazer. “Inferring friendship network structure by using mobile phone data,” in *Proceedings of the National Academy of Sciences of the United States*, 2009. [Online]. doi: 10.1073/pnas.0900282106
- [7] D. Wang, D. Pedreschi, C. Song, F. Giannotti, A. Barabási. “Human mobility, social ties, and link prediction,” in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011. [Online]. doi: 10.1145/2020408.2020581
- [8] J. Krause, S. Krause, R. Arlinghaus, I. Psorakis, S. Roberts, C. Rutz. “Reality mining of animal social systems.” *Trends in Ecology & Evolution*, vol. 28, no. 9, pp. 541-551, September 2013. [Online]. doi: 10.1016/j.tree.2013.06.002
- [9] I. Psorakis, et al. “Inferring social structure from temporal data.” *Behavioral Ecology and Sociobiology*, vol. 69, no. 5, pp. 857-866, May 2015. [Online]. doi: 10.1007/00265-015-1906-0
- [10] H. Pham, L. Hu, and C. Shahabi. “Towards integrating real-world spatiotemporal data with social networks,” in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2011. [Online]. doi: 10.1145/2093973.2094046

- [11] H. Pham, C. Shahabi, and Y. Liu. “Ebm: an entropy-based model to infer social strength from spatiotemporal data,” in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, 2013. [Online]. doi: 10.1145/2463676.2465301
- [12] H.W. Lauw, E.P. Lim, H. Pang, T.T. Tan. “Social network discovery by mining spatio-temporal events,” *Computational & Mathematical Organization Theory*, vol. 11, no. 2, pp. 97-118, July 2005. [Online]. doi: 10.1007/10588-005-3939-9
- [13] Google Earth Pro 2017. *Hong Kong 22°18'06.75"N, 114°05'29.79"E, elevation 19.18 MI*, viewed 2 November 2017.
- [14] Google Earth Pro 2017. *Spratly Islands 10°10'56.86"N, 114°13'47.33"E, elevation 823.59 MI*, viewed 2 November 2017.
- [15] Google Earth Pro 2017. *Spratly Islands 10°01'16.24"N, 115°37'16.17"E, elevation 263.6 MI*, viewed 2 November 2017.
- [16] “Group profile,” China COSCO Shipping Corporation Limited. Accessed November 7, 2017. [Online]. Available: <http://en.coscocs.com/col/col6918/index.html>
- [17] “PSC Database,” Tokyo MOU. Accessed August 20, 2017. [Online]. Available: http://www.tokyo-mou.org/inspections_detentions/psc_database.php
- [18] Google Earth Pro 2017. *Eastern Pacific 15°45'40.87"N, 117°35'15.50"E, elevation 4873 MI*, viewed 2 November 2017.
- [19] “International Convention for the Safety of Life at Sea,” International Maritime Organization. Accessed November 7, 2017. [Online]. Available: [http://www.imo.org/en/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-\(SOLAS\),-1974.aspx](http://www.imo.org/en/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-(SOLAS),-1974.aspx)
- [20] “Frequently Asked Questions,” Milltech Marine. Accessed November 7, 2017. [Online]. Available: <https://www.milltechmarine.com/faq.htm#a5>
- [21] K. Cutlip, “Spoofing: One Identity Shared by Multiple Vessels,” Global Fishing Watch, July 25, 2016. [Online]. Available: <http://blog.globalfishingwatch.org/2016/07/spoofing-one-identity-shared-by-multiple-vessels/>
- [22] “Clustering In Depth,” GitHub. Accessed November 7, 2017. [Online]. Available: <https://github.com/OpenRefine/OpenRefine/wiki/Clustering-In-Depth>
- [23] Google Earth Pro 2017. *Itu Aba Island 10°22'35.01"N, 114°21'54.37"E, elevation 17.47 MI*, viewed 2 November 2017.
- [24] Google Earth Pro 2017. *Itu Aba Island 10°22'35.01"N, 114°21'54.37"E, elevation 6778 FT*, viewed 2 November 2017.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California